STUDIES IN MICRORNA FUNCTION AND GENE DYSREGULATION IN OVARIAN CANCER

A Dissertation
Presented to
The Academic Faculty

by

Christopher G. Hill

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in the
School of Biology

Georgia Institute of Technology
December 2014

STUDIES IN MICRORNA FUNCTION AND GENE DYSREGULATION IN OVARIAN CANCER

Approved by:

Dr. John F. McDonald, Advisor
School of Biology
Georgia Institute of Technology

Dr. Eberhard Voit
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Nathan Bowen
Department of Biological Sciences
Clark Atlanta University

Dr. Jung Choi
School of Biology
Georgia Institute of Technology

Dr. I. King Jordan
School of Biology
Georgia Institute of Technology

Date Approved: November 5, 2014

For my family

ACKNOWLEDGEMENTS

I would like to thank the many people who contributed to the development and completion of this dissertation. First, my advisor, Dr. McDonald, who never let me get away with anything. I would also like to thank my committee members: Dr. Jordan, who is one of the most well-organized professors I have had; Dr. Voit, who brings clarity to even the most complex topics, as well as being a most engaging presenter; Dr. Choi, who guided my early bioinformatics course selection; and my friend, Dr. Bowen, with whom I shared many enjoyable, thought-provoking lunches. Thank you all for your patient observations, suggestions, counsel and, yes, criticisms regarding my often inchoate ideas.

Individual thanks to the current McDonald lab members, each for his/her unique contributions: Neda Jabbari, for her encouragement and for putting up with my aging memory; DeEtte Walker, the lab mom; Vinay Mittal for conversations about anything at all; Lilya Matyunina for early discussions on the scientific method and reductionism, as well as Russian chocolates; Lijuan Wang for her indefatigable cheeriness; Roman Mezencev, a friend and international political pundit; and former members Gaurav Arora, the most easy-going guy on the planet; Shubin Shahab, who told me about reduced thyroid function in post-menopausal women; Loukia Lili, who makes a delicious baklava; and Andrew Huang, a fellow computer. I thank you all for your camaraderie.

Finally, a lifetime of thanks to my wife for putting up with me and for postponing so much for too long.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

BN	Bayesian Network
cDNA	complementary DNA
CEPI	cancer epithelial cells
CIN	chromosome instability
DAG	directed acyclic graph
DAVID	database for annotation, visualization and integrated discovery
DE	differentially expressed (genes)
EOC	epithelial ovarian cancer
FBS	fetal bovine serum
FC	fold change
FDR	false discovery rate
FFL	feed-forward loop
GCRMA	Guanine Cytosine Robust Multi-Array Analysis
GEO	Gene Expression Omnibus
GO	Gene Ontology
IC	inversely correlated
LCM	laser capture micro-dissection
MAS5	MicroArray Suite 5.0
mRNA	messenger RNA
miRNA	microRNA
miR-429	MIR429

NC	no (significant) change
nt	nucleotide
OSE	ovarian surface epithelial (cells)
PC	positively correlated
RTK	receptor tyrosine kinase
SAM	statistical analysis of microarrays
SNR	signal noise ratio
TF	transcription factor
TOM	transcriptional override model

SUMMARY

Ovarian cancer, it is believed, results from the dysregulation, in normal ovarian epithelial cells, of genes responsible for the control of critical biological processes such as cell cycle and apoptosis. Since their discovery 20 years ago, microRNAs (miRNAs) have increasingly been implicated in that dysregulation due to their role fine-tuning gene expression. As a result, changes in expression levels of microRNAs have been linked with tumor growth, proliferation and metastasis. Their imputed involvement in cancer has led to the possibility of their use as biomarkers and to their potential clinical use. But the use of miRNAs as therapeutics requires accurate predictions of their direct mRNA targets and additional knowledge concerning subsequent downstream effects. Since our current apprehension of the effects of variations in miRNA expression levels is far from complete, the therapeutic value of microRNAs is somewhat limited. Using a combination of mRNA and miRNA microarray analysis to compare human gene expression between normal epithelial ovarian cells (OSE) and epithelial ovarian cancer (EOC) cells, we explored the interactions among miRNAs and genes. First, we validated *in silico* predictions of mRNA targets of microRNAs by comparing them with *in vitro* evidence after exogenous miRNA transfection. We found that miRNAs with identical seed regions shared 88% of their predicted targets, and 55% of their *in vitro* targets, confirming the importance of their 7-mer seeds. Importantly, we found that even small changes in the seed region resulted in significant changes in the set of mRNAs targeted, implying strong functional conservation of the seeds.

Next, we discovered a 3-element network motif which explains the upregulation of ~800 genes in ovarian cancer which, as predicted miRNA targets, might be expected to be down-regulated. This motif also revealed a competition, of sorts, between the repressive action of highly regulated microRNAs and the derepressive influence of down-regulated transcription factors

Finally, we developed a phenomenological network model, based on the Pearson correlation of high-throughput microarray gene expression data, to identify specific dysregulated modules involved with cell cycle and apoptosis. Our methodology identified several genes previously reported in ovarian cancer and, significantly, suggests some additional genes for further study. This technique can easily be extended to reveal dysregulated genes in other cancers.

CHAPTER 1

INTRODUCTION

Background

The notion that cancer might be better understood on the basis of genetic mutations was suggested by the early successes of the gene-centric approach to disease in the 20th century, which found, e.g., associations linking cystic fibrosis with mutations in the CFTR gene, muscular dystrophy with mutations in the dystrophin (DMD) gene, and sickle cell anemia with gene mutations in HBB (hemoglobin, beta). The discovery, in ovarian cancer, of mutations in KRAS, BRAF (Singer, Oldt et al. 2003) and other driver (and the more numerous but somewhat less pernicious passenger) genes (Haber and Settleman 2007), led to the expectation that the replacement or repair of a few miscopied bases, or the inhibition of a few key overactive RTKs (receptor tyrosine kinases, e.g., EGFR, FGFR, PDGFR) or their cognate growth factors, might halt the metastatic progress of, or at least inhibit the rapid growth of, the majority of solid cancers. But reaching, perhaps, the limits of reductionism in medicine (Ahn, Tewari et al. 2006), scientists have expanded this somewhat narrow perspective to a broader, more holistic view of biological process dysregulation, which increasingly recognizes, first, a greater variety of biological elements, from non-coding ribonucleic acids and proteins to lipids and metabolites, and second, the importance of the construction and interpretation of an interactome (Vidal, Cusick et al. 2011), a network of biological elements and their mutually influencing/interacting relationships. Among non-coding ribonucleic acids, both long (lncNRA) and short (miRNA) transcripts have become the subjects of increasing

interest. Specifically, the diagnostic, prognostic and therapeutic value of miRNAs is being evaluated (Garofalo and Croce 2011; (Pe'er and Hacohen 2011; (Gurtan and Sharp 2013).

MiRNA dysregulation

What are miRNAs, and what makes them biologically relevant in cancer? MiRNAs are short (~22-nt) non-coding RNA sequences first reported, in animals, to be involved in larval development in C. *elegans* (Lee, Feinbaum et al. 1993). Similar RNA interference had been seen in plants three years earlier (Napoli, Lemieux et al. 1990). By 2004, it was believed miRNAs might regulate up to 10% of human genes. Today, it is known that miRNAs regulate gene expression by binding and repressing mRNAs post-transcriptionally. In eukaryotes, each miRNA binds to a specific location in the 3'UTR of a targeted mRNA with a conserved 7-nt seed region (often, positions 2-8) and inhibits translation by causing mRNA destabilization and degradation (Bartel 2009). When 2 or more miRNAs have an identical seed region, they are said to belong to the same family. Since most human genes have binding sites for more than one family, in addition to multiple binding sites for each family, complex relationships emerge between cell type, miRNAs and gene targets. The resulting gene repression is reflected in the observed down-regulation of predicted miRNA targets in microarray assays.

MiRNA dysregulation has been associated with many diseases, notably cancer (Miles, Seiler et al. 2012). Thus, knowledge of the 7-nt seed sequences for the 1100 currently known (though not necessarily conserved) human miRNAs (Betel, Koppal et al. 2010), and a global census of mRNA binding sites, are critical to the successful application of

miRNAs in a clinical setting. However, since our knowledge of the downstream effects of exogenous miRNA transfection is incomplete, the therapeutic value of miRNAs is necessarily limited. We investigated two issues relating to the relevance of miRNAs in cancer.

In Chapter 2, we look at seed region variation, and confirmed the unique properties of the 7-nt (nucleotide) seed. Using observations from wet lab experiments, we verify that miRNAs with identical 7-mer seeds target a similar subset of genes. More importantly, we found that miRNAs whose seeds differ by just a single nucleotide (nt) target substantially different genes; the overlap of miRNAs with similar (but not equivalent) seeds is ~20%. Curiously, the average target overlap for two randomly chosen miRNAs is also about 20%. The fact that a single nt seed change has such a radical effect on the miRNA target population results in strong evolutionary pressure to preserve seeds, and helps explain the existence and conservation of miRNA families.

In Chapter 3, we present a novel model called TOM (Transcriptional Override Model) which accounts for the unexpected upregulation, in cancer, of many genes which were predicted targets of upregulated miRNAs. The mechanics of gene repression via miRNA inhibition anticipate an inverse relationship between (rising) miRNA expression and the (falling) expression of predicted mRNA targets. However, previous studies have found that less than 20% of computationally predicted mRNA targets are inversely related to their upregulated targeting miRNAs (Shahab, Matyunina et al. 2011). In contrast, using three computational prediction programs and the results of high throughput microarray experiments, we showed that the number of down-regulated miRNA targets in our cancer

samples were, in fact, significant. Importantly, using TOM, we were able to explain the unexpected upregulation of nearly 800 miRNA targets.

Inconsistent genes

Our work with miRNAs revealed the complexity and interconnectedness of miRNA-repressor-mRNA relationships, and suggested a network approach to better understand gene-to-gene interactions. By building phenomenological network models comparing correlations between millions of pairs of genes in normal and cancerous ovarian epithelial cells, we found that the number of highly correlated genes decreased dramatically between normal and cancer tissues, implying a loss of control in cancer. Assessing pairwise gene-to-gene relationships in a boolean-like manner, we noticed that most gene pairs maintained a similar, **consistent** relationship between normal epithelial cells and cancer tissue. Gene pairs which did not maintain a similar relationship between normal and cancer, we termed **inconsistent**. We found that inconsistent gene pairs tended to share a number of important characteristics: one of the genes was generally upregulated; both genes were enriched for the cell cycle and/or apoptotic processes; and usually, at least one member of an inconsistent gene pair was previously implicated in ovarian and/or other cancers. The complexity of these highly interconnected network modules suggests, contrary to popular opinion, that there are no master regulators of cancer. Instead, we believe there exists an orchestra of cancer-related genes, a few hundred at last count (Forbes, Bindal et al. 2011) which, when altered through mutation and interacting in small ensembles, can disrupt critical biological processes such as cell cycle and apoptosis (Kerr, Wyllie et al. 1972) the dysregulation of which is a prerequisite for tumor growth, proliferation, migration and metastasis. Cancer is in the network.

CHAPTER 2

FUNCTIONAL AND EVOLUTIONARY SIGNIFICANCE OF HUMAN MICRORNA SEED REGION MUTATIONS

Abstract

MicroRNAs have emerged in recent years as important regulators of cell function in both normal and diseased cells. miRNAs coordinately regulate large suites of target genes by mRNA silencing via degradation and/or translational inhibition. The mRNA target specificities of miRNAs in animals are primarily encoded within a 7 nt "seed region" mapping to positions 2-8 at the molecule's 5' end. We here combine computational analyses with experimental studies to explore the functional significance of sequence variation within the seed region of human miRNAs. The results indicate that a substitution of single nucleotides at any position within the seed region changes the spectrum of mRNA targets by >70%. The high functional cost of even single nucleotide changes within seed regions accounts for their high sequence conservation among miRNA families both within and between species and sheds light on the mechanisms underlying the evolution of miRNA regulatory control.

Author Summary

MicroRNAs (miRNAs) are an important component of regulatory systems that are disrupted in many human diseases including cancer. The fact that many miRNAs are sequentially highly conserved between species seems inconsistent with the fact that their regulatory targets can be highly divergent. We present here complimentary computational and experimental evidence that even a single nucleotide substitution at any position within the "seed" region of miRNAs has major functional consequences on the regulation

of hundreds of target genes. These findings account for the high sequence conservation of miRNA seed regions among and between species (stabilizing selection). Additionally, the results suggest an evolutionary model whereby miRNA regulatory evolution is driven by the relatively rapid loss and/or gain of miRNA target sequences in regulated messenger RNAs. Consistent with this evolutionary model, we show that human and mice miRNAs with identical seed sequences nevertheless display <40% overlap in their regulated genes.

Introduction

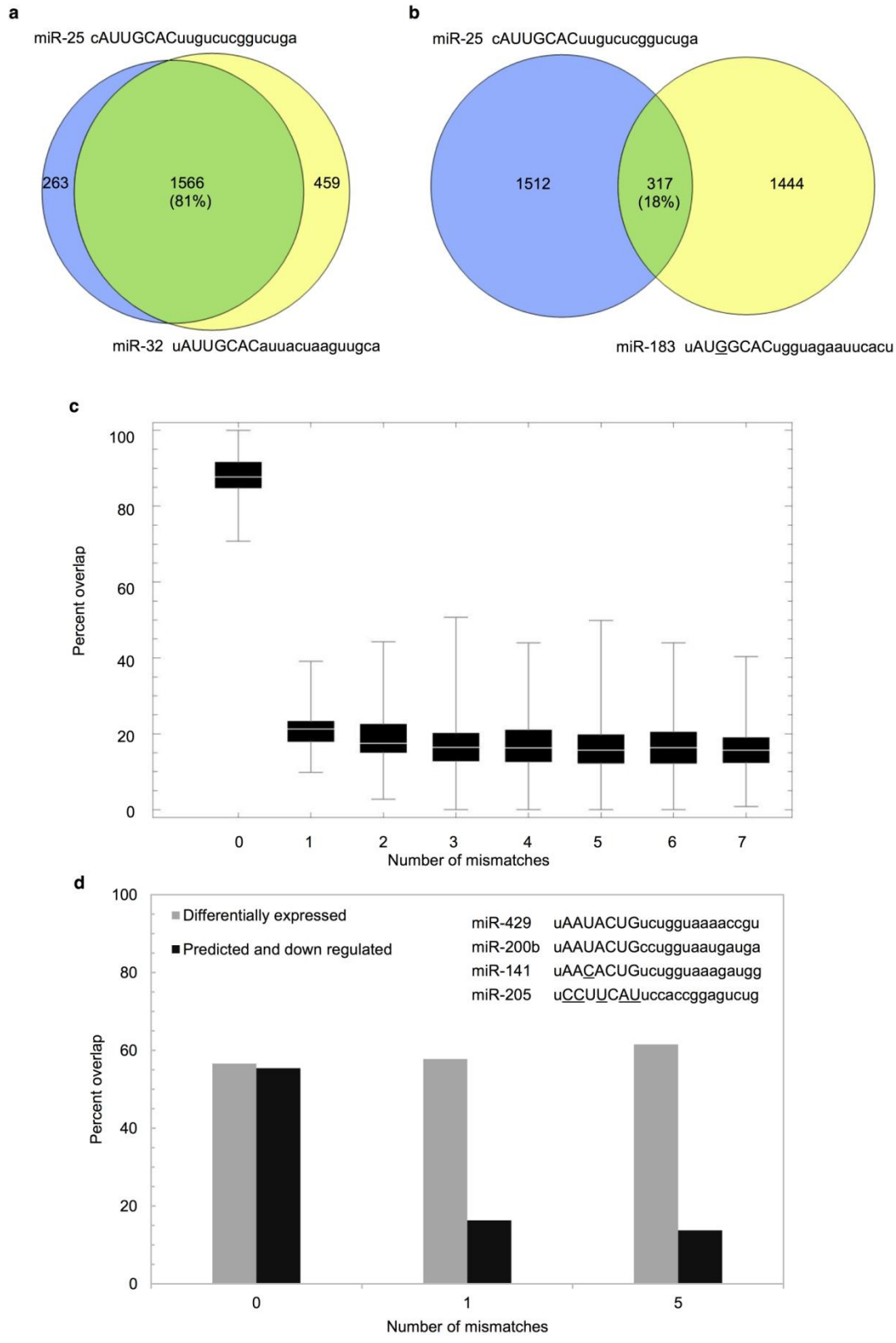
MicroRNAs (miRNAs) are small 20-22 nucleotide (nt) RNA molecules that play important regulatory roles in cell function (Nazarov, Reinsbach et al. 2013), embryonic development (Alvarez-Garcia and Miska 2005) and the onset and progression of a variety of diseases (Ha 2011), including cancer (Nana-Sinkam and Croce 2010). Like siRNAs and other small regulatory RNAs, miRNAs regulate their target genes by mRNA degradation and/or translational inhibition (Bartel 2004). However, unlike siRNAs that target one or a few genes, individual miRNAs have evolved the ability to coordinately regulate large suites of target genes, many of which may encode coordinated cellular functions (Bartel 2004; (Shahab, Matyunina et al. 2011). The mRNA target specificities of miRNAs in animals are primarily encoded within a 7 nt "seed region" mapping to positions 2-8 at the molecule's 5' end (Bartel 2009; (Wang 2014). The importance of this 7 nt sequence to miRNA function is evidenced by the fact that the seed region sequence of many miRNA families is highly conserved both within and between species (Friedman, Farh et al. 2009). Mature single-stranded miRNAs bound to the RNA-induced silencing complex (RISC) recognize their regulatory targets by Watson-Crick base

pairing to compatible sequences (usually in 3' un-translated regions or 3' UTRs) in target mRNAs.

It is estimated that there are >1800 sequentially distinct miRNAs in the human genome, each being present in a few to hundreds of copies (Kozomara and Griffiths-Jones 2014). We focused on 249 human miRNAs previously shown to be sequentially conserved across mammalian species (Lewis, Shih et al. 2003). In this study, we combine computational analyses with experimental studies to explore the functional significance of sequence variation within the seed region of human miRNAs. We find that a substitution of only a single nucleotide at any position within this 7-nt region alters the spectrum of targeted mRNAs by >70%. Further nucleotide changes have little to no additional effect. This high functional cost of even a single nucleotide change within the seed region of miRNAs explains the rigidly conserved seed sequence identity among many miRNA families both within and between species and sheds light on likely mechanisms underlying the evolution of miRNA regulatory control.

Results and Discussion

We first computed the number of nucleotide changes needed to transform one seed region into another using the Hamming distance (Hamming 1950) for each of these 249 miRNAs. We then calculated the percent overlap (cosine similarity (Tan, Steinbach et al. 2005)) of predicted targets for all pairs of these 249 miRNAs having identical seeds. For example, the percent target gene overlap for miR-25 and miR-32 (both having the seed sequence: 5'AUUGCAC) predicted by the miRanda- mirSVR (Betel, Koppal et al. 2010) algorithm is 81% (Figure 2.1a).


a
miR-25 cAUUGCACuugucucggucuga
263
1566
(81%)
459
miR-32 uAUUGCACauuacuaaguugca
b
miR-25 cAUUGCACuugucucggucuga
1512
317
(18%)
1444
miR-183 uAUGGCACugguagaauucacu
c
Percent overlap
Number of mismatches
d
Differentially expressed
Predicted and down regulated
miR-429 uAAUACUGucugguaaaaccgu
miR-200b uAAUACUGccugguaaugauga
miR-141 uAACACUGucugguaaagaugg
miR-205 uCCUUCAUuccaccggagucug
Percent overlap
Number of mismatches

.

Figure 2.1: A single nucleotide difference in the seed region of miRNAs is associated with a major change in regulated target genes. a) MiR-25 and miR-32, two miRNAs with identical seed regions (upper-case letters), have 81% overlap in their predicted target genes; b) MiR-25 and miR-183, two miRNAs with a single nucleotide difference within their respective seed regions have only 18% overlap in their predicted target genes; c) The overlap of predicted targets for all 249 pairs of conserved miRNAs grouped by the number of mismatches in their respective seed regions. The computed percent overlaps (cosine similarity) are presented as box and whisker plots [the bottom and top of each box represent the first and third quartiles of variation while the band inside the box represents the median (second quartile) value*;* the *"*whiskers" represent variability outside the upper and lower quartiles]; d) Four members of the miR-200 family (miR-141, miR-200b, miR-205, miR-429) and a negative control were independently transfected into HEY cells and, after 48 hrs, RNA was extracted and subjected to gene expression analysis (Affymetrix, U133). The histogram displays the observed overlap of all significantly differentially expressed genes (light shading) between miR-429 transfected cells and cells transfected with miR-200b (identical seed region with miR429), miR-141 (1 seed region mismatch with miR-429) or miR-205 (5 seed region mismatches with miR-429). Shown also (dark shading) is the percent overlap between significantly down-regulated mRNAs that are predicted targets of the transfected miRNAs.

The median percent overlap between all 249 pairs of conserved miRNAs with identical seeds is 88% (Figure 2.1c). We next independently computed the percent overlap of predicted mRNA targets for pairs of miRNAs having seeds that differed by 1 to 7 nucleotides. The results (Figure 2.1b, 2.1c) demonstrate that even a single nucleotide mismatch in the seed regions of two miRNAs is predicted to reduce the percent overlap among their respective targeted mRNAs by >70%. The generality of these predictions was corroborated independently by conducting the same analyses using two additional target prediction algorithms (TargetScan (Friedman, Farh et al. 2009) and PicTar (Krek, Grun et al. 2005)).

To experimentally test these predictions, we selected members of the miR-200 family of human miRNAs that share identical seed regions (miR-429 vs. miR-200b), differ by a

single nucleotide (miR-429 vs. miR-141) or differ by multiple nucleotides (e.g. 5 nucleotide differences: miR-429 vs. miR-205) within their respective seed regions (Figure 2.1d). These miRNAs, as well as, a synthetic control miRNA with no sequence homology to any known miRNA were independently transfected into the well characterized ovarian cancer HEY cell line (Buick, Pullano et al. 1985) and, after 48 hrs, RNA was extracted and subjected to gene expression analysis (Affymetrix, U133) as previously described (Shahab, Matyunina et al. 2012; (Jabbari, Reavis et al. 2014) (see Tables A.1, A.2, A.3 and A.4 for detailed results of microarray analyses).

Ectopic over-expression of miRNAs in cells induces both direct and indirect changes in gene expression (Shahab, Matyunina et al. 2011) while target prediction algorithms only predict direct interactions, *i.e*. mRNAs predicted to be directly targeted by specific miRNAs. Thus, from our experimental dataset, we were interested in comparing those down-regulated mRNAs predicted by miRanda-mirSVR to be targeted by the miRNAs employed in our transfection experiments. As in the computational analyses, we computed the percent overlap between genes significantly down-regulated in response to transfection of miRNAs having identical seeds, seeds with a single nucleotide difference and seeds with 5 nucleotide differences (Figure 2.1d). The remarkable concordance between the computationally predicted and experimentally observed results corroborates the hypothesis that even a single nucleotide difference within the seed region of miRNAs is associated with substantial functional cost.

We were next interested in determining if the position of single nucleotide changes within the seed region of miRNAs has a significant effect. Drawing again on the miRanda-mirSVR target predictions of the 249 sequentially conserved miRNAs, we divided pairs of miRNAs whose seeds differed by a single nucleotide into separate groups based on the position (positions 2-8) of the mismatch within their respective seed regions. The results presented in Figure 2.2a display the percent overlap of predicted mRNA targets among miRNAs differing by a single nucleotide at different positions within their respective seed regions. The results indicate that a mismatch at any position within the seed region is predicted to result in the same significant reduction in overlap relative to a perfect 7 nt match.

To experimentally test the effect of the position of single nucleotide differences within seed regions on targeted mRNAs, we constructed a synthetic miRNA identical in seed sequence to miR-429 (and miR-200b) but for a single nucleotide change at position 2 (M12). The naturally occurring miR-141 is also identical in seed sequence to miR-429 but for a single nucleotide change at position 4 (Figure 2.2b). These miRNAs (and the negative control) were independently transfected into HEY cells and, after 48 hrs, RNA was extracted and subjected to gene expression analysis (Affymetrix, U133). The results presented in Figure 2.2b (see Tables A.1, A.2, A.3 and A.5 for detailed microarray results) confirm that the position of the mismatch among the miRNAs tested had no significant effect.

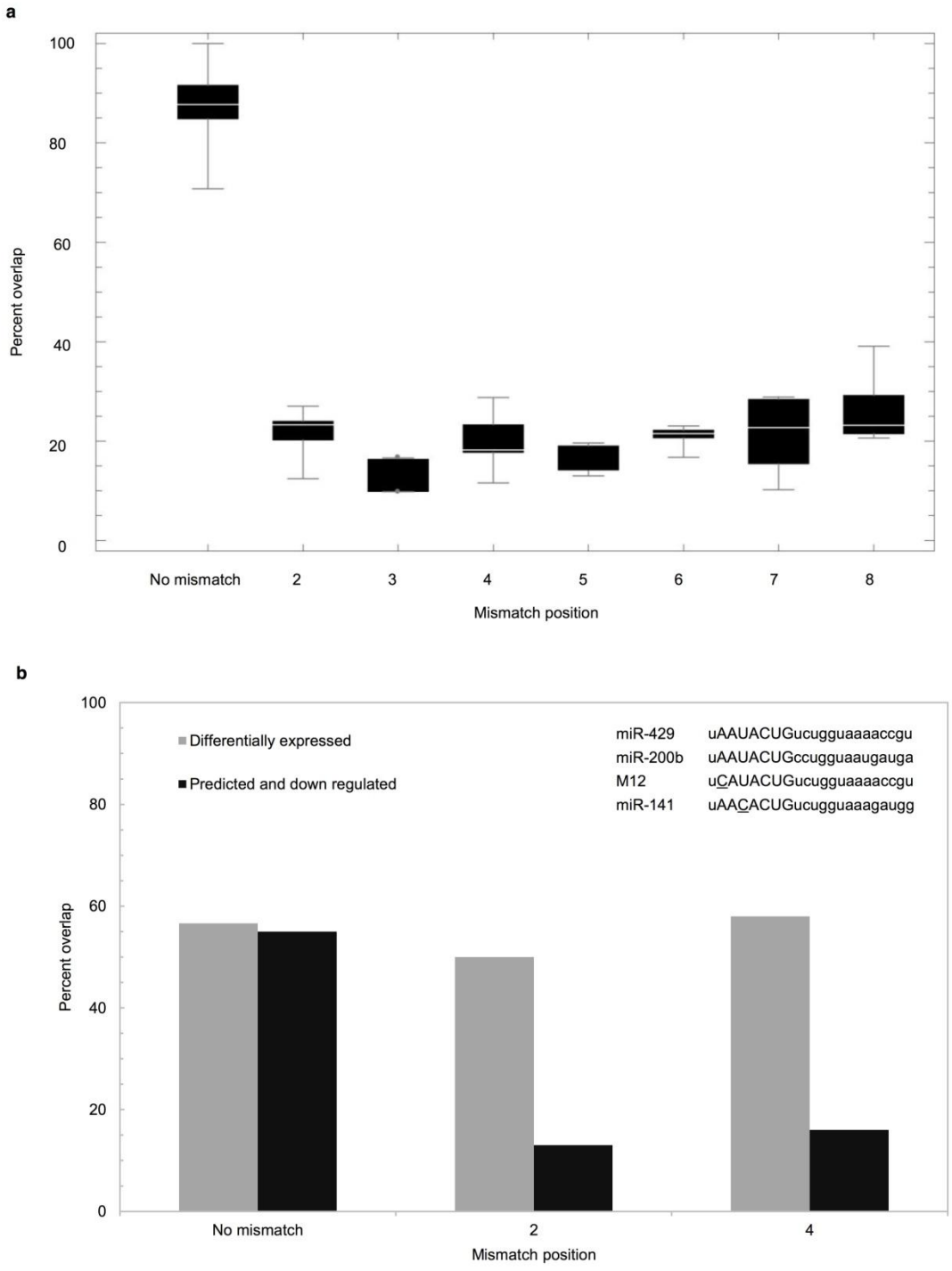


Figure 2.2: Single nucleotide changes at any position within the seed region of miRNAs are associated with a major change in regulated target genes. a) Shown is the percent overlap (cosine similarity) of predicted mRNA targets among 30,876 pairs of miRNAs differing by a single nucleotide at different positions within their respective seed regions. b) mir-429, miR-200b, miR141 and a synthetic miRNA identical in sequence to miR-429 but for a single nucleotide change at position 2 (M12) (and negative

control) were independently transfected into HEY cells. The histogram displays the observed overlap of all significantly differentially expressed genes (light shading) between miR-429 transfected cells and cells transfected with miR-200b (identical seed region with miR429), miR-141 (1 seed region mismatch at position 4) and M12 (1 seed mismatch at position 2). Shown also (dark shading) is the percent overlap between significantly down-regulated mRNAs that are predicted targets of the transfected miRNAs.

The above results indicate that even a single nucleotide substitution within the seed regions of miRNAs results in large changes in target mRNAs and consequentially altered global patterns of gene expression. These findings further imply an evolutionary model whereby strong stabilizing selection is maintaining rigid conservation of miRNA seed sequences both within and between species. Individual target genes, on the other hand, may acquire and/or lose miRNA regulatory control(s) through even single nucleotide substitutions in miRNA target sequences complimentary to miRNA seeds (typically within 3' UTRs) (Figure 2.3). Any functional consequence of such mutations would be incurred on the individual gene level rather than on the multi-gene level associated with miRNA seed region mutations. Thus, the model predicts that although seed regions may be highly conserved both within and between species due to strong stabilizing selection, the spectrum of genes regulated by these conserved miRNAs may vary significantly, especially between more distantly related species where there has been ample time/opportunity for individual genes to acquire variations in their target sequence(s) and to re-associate themselves with other, presumably adaptive, miRNA regulatory controls (directional selection).

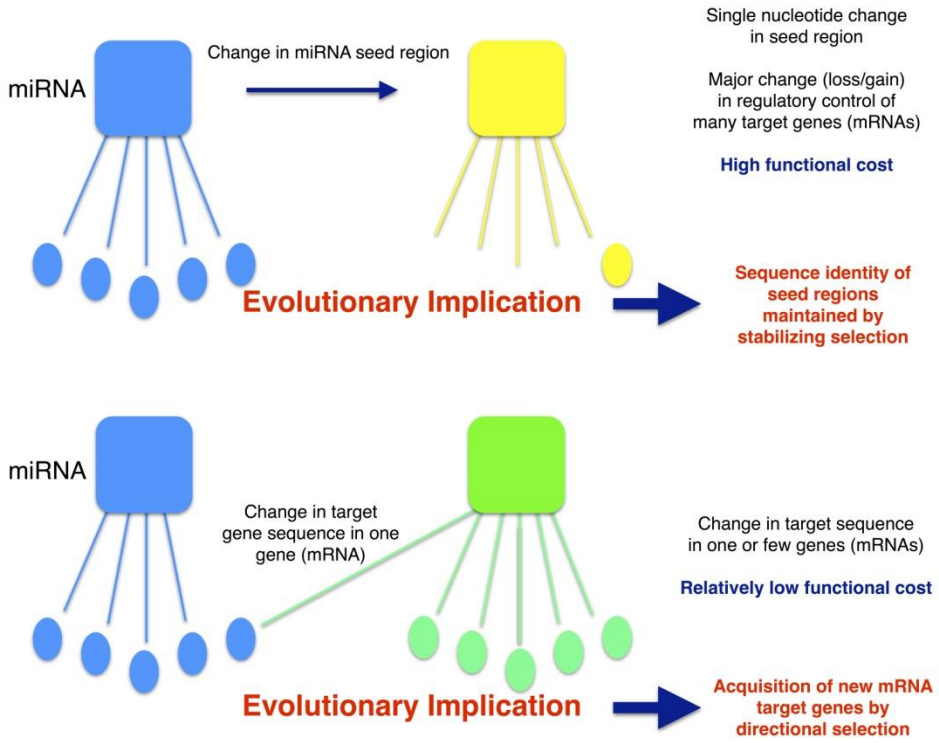


Figure 2.3: The high functional cost of even single nucleotide changes in miRNA seed regions implies a mechanism of miRNA regulatory evolution.

To test this prediction, we selected two miRNAs (miR-429 and miR-200b) that have identical seed regions in both humans and mice (Figure 2.4a). We again employed the miRanda-mirSVR algorithm to predict the respective orthologous mRNA targets of these two miRNAs (human: hsa-miR-429, hsa-miR-200b; mouse: mmu-miR-429, mmu-miR-200b) in both species. As shown in Figure 2.4b, the percent overlap between the predicted gene targets of these two miRNAs (intra-specific) is >90% (mouse: 93.3%; humans: 91.8%) in both species. However, despite the fact that the human and mouse miRNAs share sequentially identical seed regions, they display < 40% overlap among their respective target genes/mRNAs in the non-native species (inter-specific) (Figure

2.4b). To determine if these differences are representative of other sequentially conserved miRNAs, we computed the percent overlap of genes targeted by the 249 miRNAs sequentially conserved in mouse and humans. The results confirm that the average overlap between targeted genes in mouse and humans is < 30% (Figure 2.4c). These results are consistent with the hypothesis that while miRNA seed regions may be selectively conserved across species, target genes maintain relative flexibility to acquire or lose miRNA regulatory controls by even single nucleotide changes within their respective miRNA target sequences (typically within 3' UTRs).

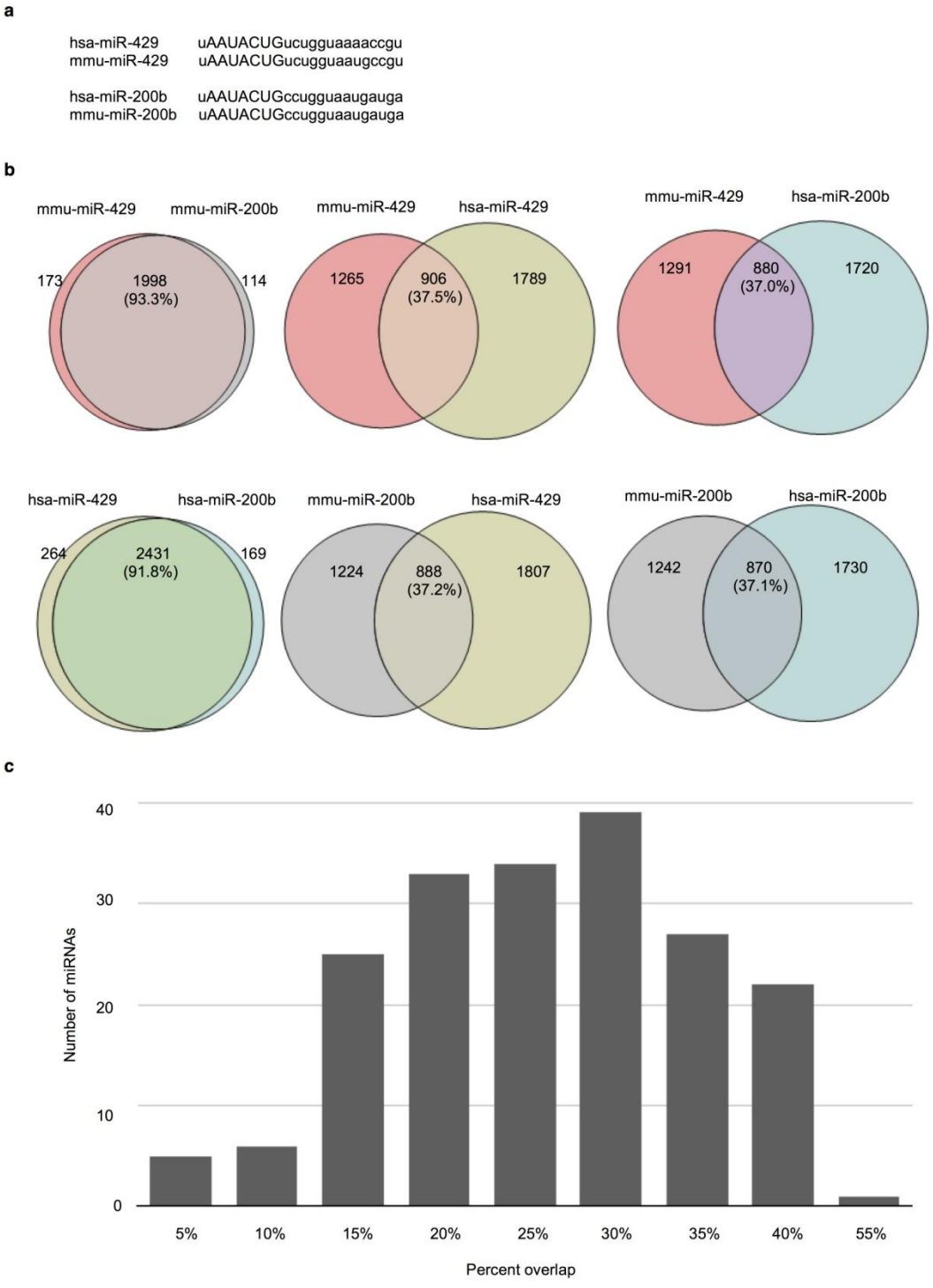


Figure 2. 4: Although mouse and human miRNAs share sequentially identical seed regions, they regulate highly divergent groups of target genes. a) Sequence alignment between human (has-miR-429/200b) and mouse (mmu-miR-429/200b) miR-429 and miR-200b miRNAs. Despite the evolutionary distances between these two species, the respective seed regions are sequentially identical, reflective of strong stabilizing

selection. b) Venn diagrams showing the % overlap between human and mouse orthologs that are predicted targets (miRanda-mirSVR) of human and mouse miR-429 and miR-200b. c) The average percent overlap of genes targeted by all 249 sequentially conserved mouse and human miRNAs is < 30%. The results are consistent with a model of directional selection operating on variation in miRNA target sequences (typically mapping to 3' UTRs) to re-position genes under optimally adaptive miRNA regulatory control(s).

Collectively, our findings indicate that miRNAs initially evolve to regulate large suites of target genes. Thereafter, the sequential integrity of miRNA seed regions is maintained by strong stabilizing selection due to the high functional cost of even a single nucleotide mutation at any position within the miRNA. In contrast, nucleotide mutations in the target sequences of individual genes, being of substantially lower functional cost, allow for a relatively rapid repositioning of miRNA-target gene associations. Indeed, a variety of scenarios might arise to buffer the possible negative effects of target sequence mutations in regulated genes. For example, duplication of specific target sequences within regulated genes could serve to mask the impact of the sudden loss of existing miRNA regulatory controls while permitting genes to explore the potential adaptive benefits of acquiring new miRNA regulatory controls.

Materials and Methods

Cell culture and transfections

HEY ovarian cancer cells, provided by Gordon B. Mills (MD Anderson Cancer Center, Houston, TX), were cultured in RPMI-1640 (Mediatech, Manassas, VA) with 10% Fetal Bovine Serum (FBS, *Atlanta Biologicals, GA*) and 1% antibiotic-antimycotic solution (Mediatech-Cellgro, Manassas, VA) and incubated at 37°C and 5% CO_2. The transfection

protocol was as described described (Shahab, Matyunina et al. 2012; (Jabbari, Reavis et al. 2014). Briefly, triplicate wells of exponentially growing cells were transfected with miR-141, miR-200b, miR-429, miR-205 and M12 miRNAs purchased as Pre-miR™ miRNA Precursors (Ambion, Austin, TX). Transfections were performed using Lipofectamine 2000 transfection reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Ambion Pre-miR™ miRNA Precursor Negative Control was used as a negative control.

miRNA target prediction

Predicted miRNA targets (mRNAs) were established using the miRanda-mirSVR (Betel, Koppal et al. 2010) target prediction algorithm (mirSVR score ≤-0.2). Corroborative predictions were carried out using the Targetscan (Friedman, Farh et al. 2009) and PicTar (Krek, Grun et al. 2005) prediction algorithms.

RNA isolation and whole genome microarray

RNA was extracted from transfected cells using RNeasy Mini RNA isolation kit (QIAGEN, Valencia, CA). Microarray experiments were performed as previously described (Shahab et al. 2012). Briefly, RNA samples with high integrity were converted to cDNA and amplified with Applause 3'-Amp System (NuGen, San Carlos, CA). The cDNA was fragmented and Biotin labeled using the Encode Biotin Module (NuGen). Labeled cDNA was then hybridized to Affymetrix HG-U133 Plus 2.0 arrays and analyzed with GeneChip Scanner 3000 (Affymetrix, Santa Clara, CA). All microarray

data are deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov) SuperSeries number GSE56973.

Microarray data analysis

Determination of differential expression of genes after transfection was carried out as follows: 1) normalization by *Microarray Suite 5.0* (MAS5.0) using the Affymetrix Expression Console v1.1 to obtain present/absent calls. Probesets with less than 50% present calls across all samples were removed; 2) quality control and normalization by GC *Robust Multi-array Average (*GCRMA) as implemented by Array Analysis (Eijssen, Jaillard et al. 2013) to obtain expression values for experimental and negative controls. Using only those CEL files that passed quality controls, we calculated mean signal values for triplicate samples for miR-429, miR-200b, miR-205 and M12 (duplicate for miR-141) and negative control. The log ratio representing the change in expression between experimental miRNA and negative control transfections was calculated; 3) calculation of signal-to-noise ratio (SNR = mean/standard deviation) was used to select the probe set with the highest SNR. This guarantees the highest signal change while minimizing noise; and 4) multiple testing correction was done using SAM (Significance Analysis of Microarrays (Tusher, Tibshirani et al. 2001)) with a false discovery rate (FDR) < 2%.

miRNA target overlap

Target overlap between miRNAs was determined using cosine similarity (Tan, Steinbach et al. 2005) and as follows: the total number of overlapping genes was divided by the square root of the product of the total number of genes targeted by both miRNAs.

Orthologous genes

Orthologous genes in human and mouse were extracted using the BioMart data-mining tool (Kasprzyk 2011).

CHAPTER 3

TRANSCRIPTIONAL OVERRIDE: A REGULATORY NETWORK MODEL OF INDIRECT RESPONSES TO MODULATIONS IN MICRORNA EXPRESSION

Abstract

Background

MicroRNAs (miRNA) are small non-coding regulatory molecules which have been implicated in a variety of diseases including cancer. This suggests their usefulness not only as early indicators of disease, but as a potential new class of therapeutics. A significant hurdle to their clinical application is our inability to predict the response, at both genomic and cellular levels, to the repression of miRNA targets. So while the mRNA binding sites of the targets of individual miRNAs can be computationally identified, a complete understanding of the indirect molecular effects of perturbations in miRNA levels remains a major challenge in cancer systems biology.

Results

We describe and provide experimental support for a network model to explain the unanticipated upregulation of predicted targets of highly expressed microRNAs. We found 31 microRNAs upregulated in ovarian cancer samples; many of their miRanda-mirSVR predicted targets were, as measured by mRNA levels, down-regulated. However, over 1500 putative targets were upregulated in our data. We noticed that the expression

levels of hundreds of transcriptional repressors were negatively correlated with predicted targets of miRNAs in normal tissue. That is, variations across normal ovarian samples revealed an inverse relationship between the expression of miRNA targets and transcriptional repressors. We show that microRNAs, their mRNA targets, and down-regulated transcriptional repressor genes can form incoherent feed-forward loops (FFLs) which result in the unanticipated upregulation of those targets in cancer (Figure 3.1A). We call this FFL motif the Transcriptional Override Model (TOM).

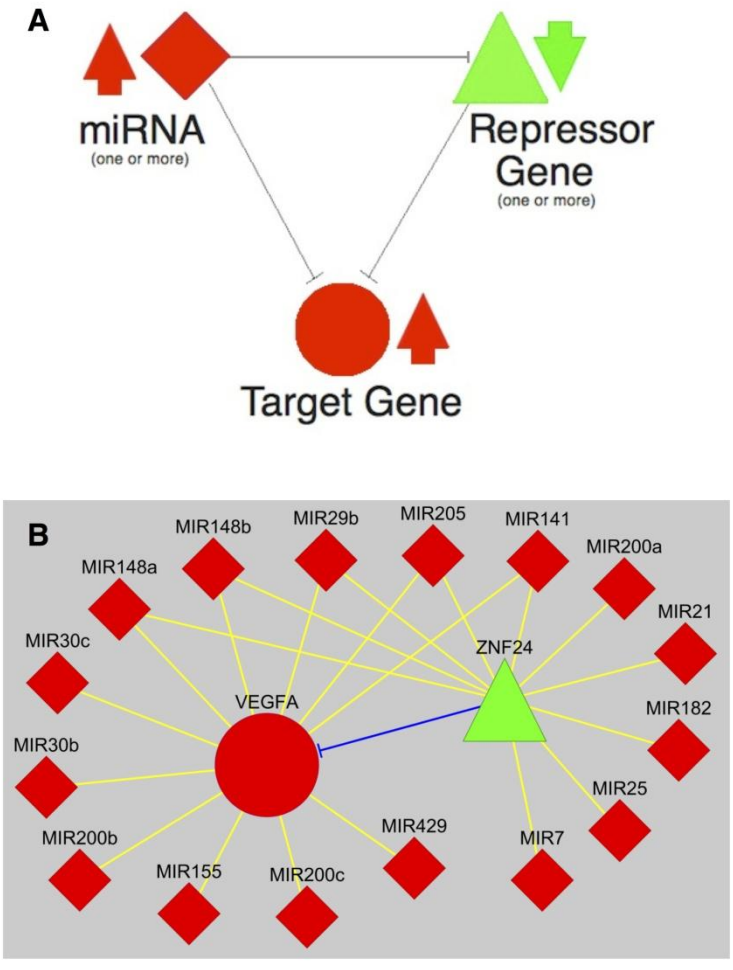


Figure 3.1: The transcriptional override model (TOM). This network motif A) typifies how the expected down-regulation of target genes by elevated levels of regulating miRNAs may be masked or overridden by target gene derepression mediated by miRNA-

induced downregulation of repressor genes. B) MiRNA-mediated derepression of repressor gene *ZNF24* (green triangle) overrides the expected downregulatory effects of miRNAs (red diamonds) on VEGFA expression (red circle).

Conclusions

The regulatory effects of changes in miRNA expression levels *in vivo* are often indirect and complex, but may be amenable to systems level modeling. Although TOM was developed and validated in the context of ovarian cancer, we believe it may contribute to a better understanding of critical genomic interactions in other diseases, and to the use of miRNA-based strategies for the treatment of cancer.

Background

Human miRNAs regulate gene expression post-transcriptionally by degrading target mRNAs and/or blocking their translation (Bartel 2009). As a consequence, mRNA expression changes are expected to be inversely correlated (IC) with changes in levels of their targeting miRNAs. Although this expectation has been validated in studies of individual miRNAs and specific mRNA targets, the expected inverse relationship is often not observed in global transcriptome level studies (Lim, Lau et al. 2005; (Shahab, Matyunina et al. 2011; (Matkovich, Hu et al. 2013). While these unexpected findings may, in some instances, be attributable to inaccuracies in miRNA target prediction algorithms (Witkos, Koscianska et al. 2011) or translational repression (Huntzinger and Izaurralde 2011), recent evidence suggests that much of the unexpected regulatory effects may be the result of feed-back or feed-forward loops (FFLs) and/or other system level complexities (Tsang, Zhu et al. 2007; (Shahab, Matyunina et al. 2011). Seeking an explanation for unexpected interactions in ovarian cancer, we probed our data for the

existence of an FFL consisting of an upregulated miRNA, a down-regulated transcriptional repressor, and a mutual gene target. That gene target must satisfy 2 criteria: 1) it must be a predicted miRanda-mirSVR target of one or more of the 31 upregulated miRNAs observed in ovarian cancer, and 2) it must be negatively correlated with 1 or more known transcriptional repressors in normal tissue. We show that that FFL motif, TOM, accounts for 865 upregulated miRNA targets.

Results

To systematically investigate this model, we employed microarray gene expression profiling to compare differences in expression levels of mRNAs and miRNAs in ovarian surface epithelial cells (OSE) vs. serous papillary epithelial ovarian cancer (EOC) cells isolated from patient tissues (Additional file 1) by laser capture microdissection. Gene expression profiling identified 5910 significantly differentially expressed genes (mRNAs) between OSE and EOC (Additional file 2). Of these, 2232 (38%) were significantly upregulated and 3678 (62%) downregulated in EOC. MiRNA expression profiling identified 31 significantly differentially expressed miRNAs between OSE and EOC. All of these miRNAs were upregulated in EOC (Table 3.1).

Table 3.1: MicroRNAs upregulated in cancer. MiRNA expression profiling identified 31 significantly upregulated miRNAs between OSE and EOC.

microRNA	p-value	fold-change
MIR7	0.00004	25.81
MIR10a	0.00305	28.25
MIR17	0.01120	14.12
MIR18a	0.00001	38.05
MIR18b	0.00000	46.21
MIR19a	0.00001	41.64
MIR19b	0.00103	61.39
MIR20a	0.00707	14.42
MIR20b	0.03363	19.16
MIR21	0.01296	14.12
MIR25	0.01993	5.58
MIR29b	0.00874	23.10
MIR30b	0.01763	11.55
MIR30c	0.01292	26.54
MIR93	0.02678	9.58
MIR106a	0.00505	20.53
MIR106b	0.00204	32.90
MIR128	0.00542	15.78
MIR130b	0.03129	11.71
MIR141	0.00000	118.60
MIR143	0.00792	5.98
MIR148a	0.00406	10.70
MIR148b	0.01725	7.94
MIR155	0.00179	35.75
MIR181d	0.02989	8.40
MIR182	0.00046	58.49
MIR200a	0.00264	39.12
MIR200b	0.01871	14.52
MIR200c	0.00232	13.36
MIR205	0.00018	67.18
MIR429	0.00181	21.11

Employing three commonly used miRNA target prediction algorithms (miRanda-mirSVR (Betel, Koppal et al. 2010), TargetScan (Friedman, Farh et al. 2009), SVMicrO (Liu, Yue et al. 2010)), we identified putative mRNA targets of these 31 miRNAs to determine if differences in their levels of expression between the OSE and EOC samples were IC,

positively correlated (PC) or unchanged (NC). Based on the established molecular mechanism of miRNA regulation, levels of miRNAs are expected to be IC with levels of their target mRNAs. Contrary to this expectation and consistent with previous findings [3], we observed a consistently low percentage (23-31%) of target mRNAs displaying expression level changes IC with their regulating miRNAs (Table 3.2). For the remainder of this study, we relied on the predictions from miRanda-mirSVR.

Table 3.2: Comparison of microRNA target prediction algorithms. This table shows the percentage of miRNA target genes displaying changes in expression between OSE and EOC as predicted by three popular target prediction algorithms, miRanda/mirSVR, TargetScan and SVMicrO.

Algorithm	Inversely Correlated	Positively Correlated	No Change Detected	Total
miRanda/mirSVR	24% (3110)	13% (1719)	62% (8015)	12844
TargetScan	31% (1519)	16% (779)	54% (2681)	4979
SVMicrO	23% (2027)	13% (1164)	64% (5787)	8978

The transcriptional override model (TOM) postulates that downregulation of target genes induced by elevated levels of regulating miRNAs may be masked (NC) or overridden (PC) by increases in expression mediated by the downregulation of repressor genes that are themselves targets of upregulated miRNAs (Figure 3.1A). The possibility of such feed-forward loops was prompted by the fact that several of the predicted mRNA targets of the 31 overexpressed miRNAs encode documented repressors of gene expression (*e.g.*,

ZNF24 (Jia, Hasso et al. 2013), *YY1* (Bonavida and Baritaki 2011), *SPEN* (Ariyoshi and Schwabe 2003), *BACH1* (Kitamuro, Takahashi et al. 2003)). MiRNA-mediated downregulation of these repressor genes would be expected to result in the derepression of their respective target genes and a consequent increase in levels of expression. If these derepressed gene targets were also the targets of upregulated miRNAs, the expected downregulation of these genes by the miRNAs (IC) could be masked (NC) or overridden (PC). For example (Figure 3.1B), one of the predicted targets of ten of the 31 miRNAs upregulated in cancer is the well-documented repressor gene *ZNF24* (Jia, Hasso et al. 2013). Consistent with the fact that *ZNF24* is targeted by upregulated miRNAs, its expression in EOC is significantly reduced. An experimentally validated target of *ZNF24* is *VEGFA* (Jia, Hasso et al. 2013). Despite the fact that *VEGFA* is itself directly targeted by 11 upregulated miRNAs (including five of those targeting *ZNF24*), its level of expression is significantly increased (PC) in EOC. These results are consistent with the hypothesis that *ZNF24*-mediated derepression is overriding the expected downregulatory effects of the upregulated miRNAs on *VEGFA* expression.

Although many of the genes falling within the NC category could be the result of partial transcriptional override, they might also simply be the result of no or slight miRNA regulatory effects. Since we cannot experimentally distinguish between these two possibilities, we will operationally only consider PC differences in expression as being inconsistent with the expected IC differences.

To evaluate TOM further, we identified 105 genes that are 1) targets of one or more of the 31 upregulated miRNAs, 2) significantly downregulated in our cancer samples and 3)

previously characterized as transcriptional repressors (Ashburner, Ball et al. 2000) (Table B.3). The targets of ten (Table 3.3) of these 105 genes have been previously identified in the Transcription Factor Binding Site database (TRANSFAC) (Matys, Kel-Margoulis et al. 2006).

Table 3.3: Ten down-regulated transcriptional repressors. Ten genes characterized as validated repressors and predicted targets of one or more of the 31 miRNAs upregulated in EOC.

Probeset_ID	Gene_Symbol	p-value	fold-change
204194_at	BACH1	0.00056	-3.65
221234_s_at	BACH2	0.00033	-2.17
209271_at	BPTF	0.00405	-2.20
225572_at	CREB1	0.00156	-1.98
224889_at	FOXO3	0.00390	-2.04
210002_at	GATA6	0.00000	-56.60
208328_s_at	MEF2A	0.02267	-2.16
209239_at	NFKB1	0.00116	-1.66
209706_at	NKX3-1	0.00000	-20.57
201901_s_at	YY1	0.01216	-2.70

This resulted in 843 genes (Additional file 4) predicted to be directly targeted by both the ten downregulated repressors and one or more of the 31 upregulated miRNAs. From the perspective of miRNA regulation, all 843 of these target genes are expected to be downregulated while from the perspective of the downregulated repressor genes all the targets are expected to be upregulated. The observed reality lies somewhere between

these two expectations (Figure 3.2). TOM predicts that the response of any particular target gene will be determined by the relative strengths of these two opposing regulatory controls.

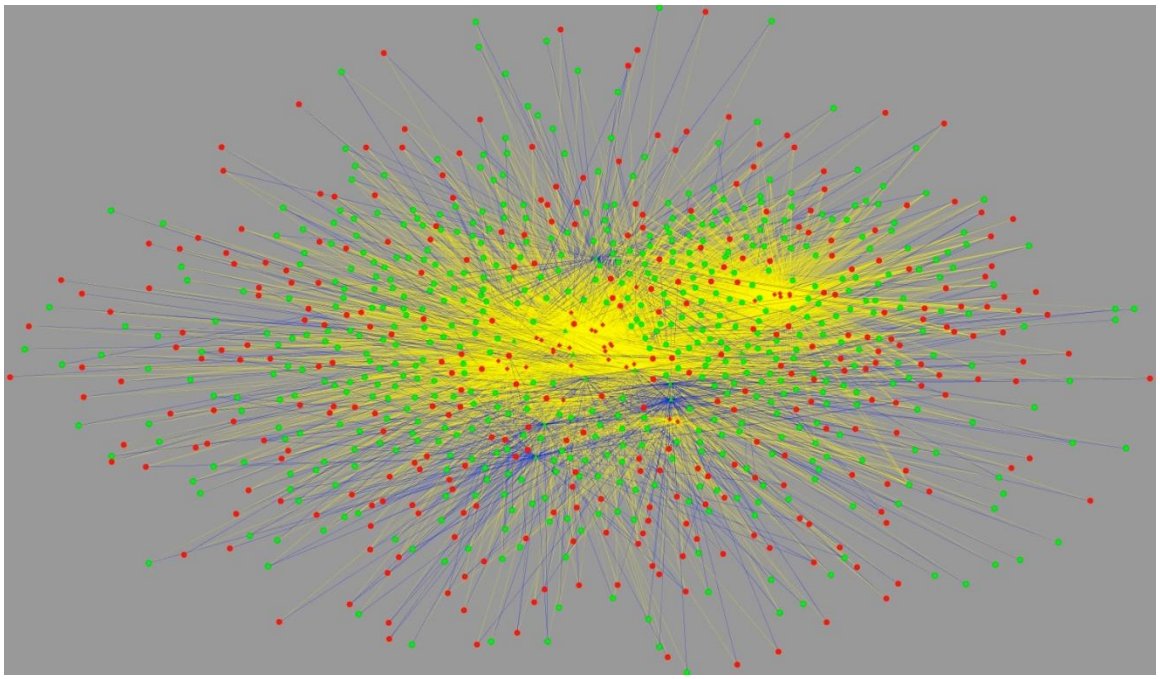


Figure 3.2: Highly interconnected network of 31 microRNAs, repressors and their mutual targets. Relationships among 31 miRNAs upregulated in cancer (red diamonds), ten downregulated repressors that are targets of one or more of the upregulated miRNAs (green triangles) and 843 genes that are the gene targets (red and green circles) of both the upregulated miRNAs and the downregulated repressor genes. Black lines depict the regulatory connection between miRNAs and their target genes; blue lines depict the regulatory connection between repressors and target genes. From the perspective of the upregulated miRNAs, all target genes should be downregulated (green); from the perspective of the downregulated repressors, all target genes should be derepressed (upregulated) (red). According to TOM, the response of any particular target gene will be determined by the relative strengths of these opposing regulatory controls.

The number of miRNAs targeting individual human genes (mRNAs) is known to vary from zero to over 100 with an estimated average of 7.1 miRNA targets per gene (John,

Enright et al. 2004). Thus, the relative strength of the regulatory effect of miRNAs on target genes might be expected to be a function of the number of miRNAs targeting individual genes. To explore this possibility, we grouped the 843 predicted gene targets of the 31 upregulated miRNAs and the 10 downregulated repressors into 5 groups based upon the number of upregulated miRNAs predicted to target each gene (Figure 3.3A). A sixth group was comprised of targets of the ten repressor genes (TRANSFAC) but not predicted targets of any of the 31 miRNAs. For each group, we computed the percentage of target genes that were upregulated (PC). In addition, for each group we divided the target genes into those predicted to be regulated by a single repressor vs. those predicted to be regulated by multiple repressors (Figure 3.3B).

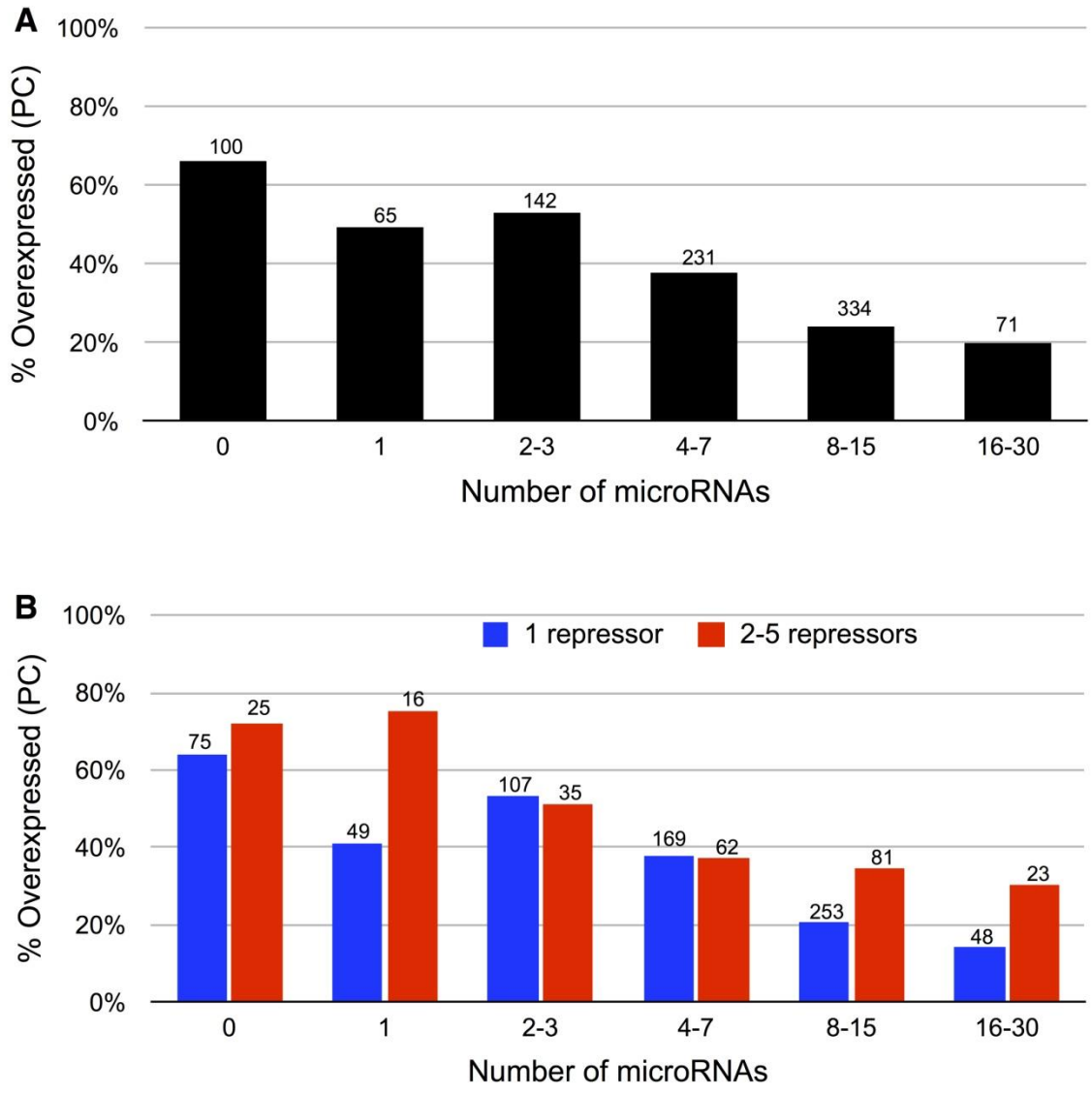


Figure 3.3: Analysis of changes in expression of miRNA regulated repressor genes and their predicted target genes is consistent with TOM. The predicted target genes (843) of 31 upregulated miRNAs and 10 downregulated repressors were divided into 6 groups based upon the number of upregulated miRNAs targeting each gene. (The number of genes included in each group is shown over the bars.) A) The strength of miRNA regulatory control on target genes increases with the number of miRNAs targeting individual genes. Bars represent the percentage of upregulated genes. The fact that ~20% of genes targeted by even large numbers (16-30) of upregulated miRNAs continue to display the unexpected PC is consistent with TOM. The chi-square test for trend is ($X^2=67.34; p < 0.0001$). B) Genes targeted by 2-5 downregulated repressors (red bars) override miRNA regulatory effects on co-regulated target genes relative to genes targeted by one downregulated repressors (blue bars). The chi-square test for trend for the blue bars this figure (3.3B) is ($X^2=5.25; p < 0.0219$).

The results demonstrate that as the number of upregulated miRNAs targeting individual genes increases, the percentage of target genes displaying the unexpected PC change in expression decreases. These results are consistent with TOM and indicate that as the relative strength of the miRNA regulatory effect increases, the impact of the opposing derepression effect mediated by the downregulated repressor genes is diminished. However, the fact that ~20% of genes targeted by even large numbers (>15) of upregulated miRNAs continue to display the unexpected PC indicates that, in some cases, the magnitude of derepression is sufficient to completely override miRNA regulation. The results presented in Figure 3.3B suggest that genes targeted by multiple repressors tend to be associated with a higher percentage of PC genes than those targeted by a single repressor. The effect, however, is not as consistent as observed with increasing numbers of regulating miRNA likely due to the relatively low number of repressor genes in this dataset and the fact that not all repressor genes can be expected to exert the same magnitude of regulatory control.

We were next interested to see if the model's ability to account for trends observed using the limited dataset described above might also extend more globally. We divided all differentially expressed genes including those that are predicted gene targets of the 31 upregulated miRNAs (4829) and those that are not (1081), into 6 groups based on the number of miRNAs targeting each gene. The results (Figure 3.4A) demonstrate a clear inverse relationship between the number of miRNAs targeting genes and the percentage of these genes displaying the unexpected PC. Again, however, we found that ~20% of

genes targeted by even large numbers (>15) of upregulated miRNAs continue to display the unexpected PC consistent with the hypothesis that the magnitude of repressor gene mediated derepression is, in some instances, sufficient to completely override miRNA regulation.

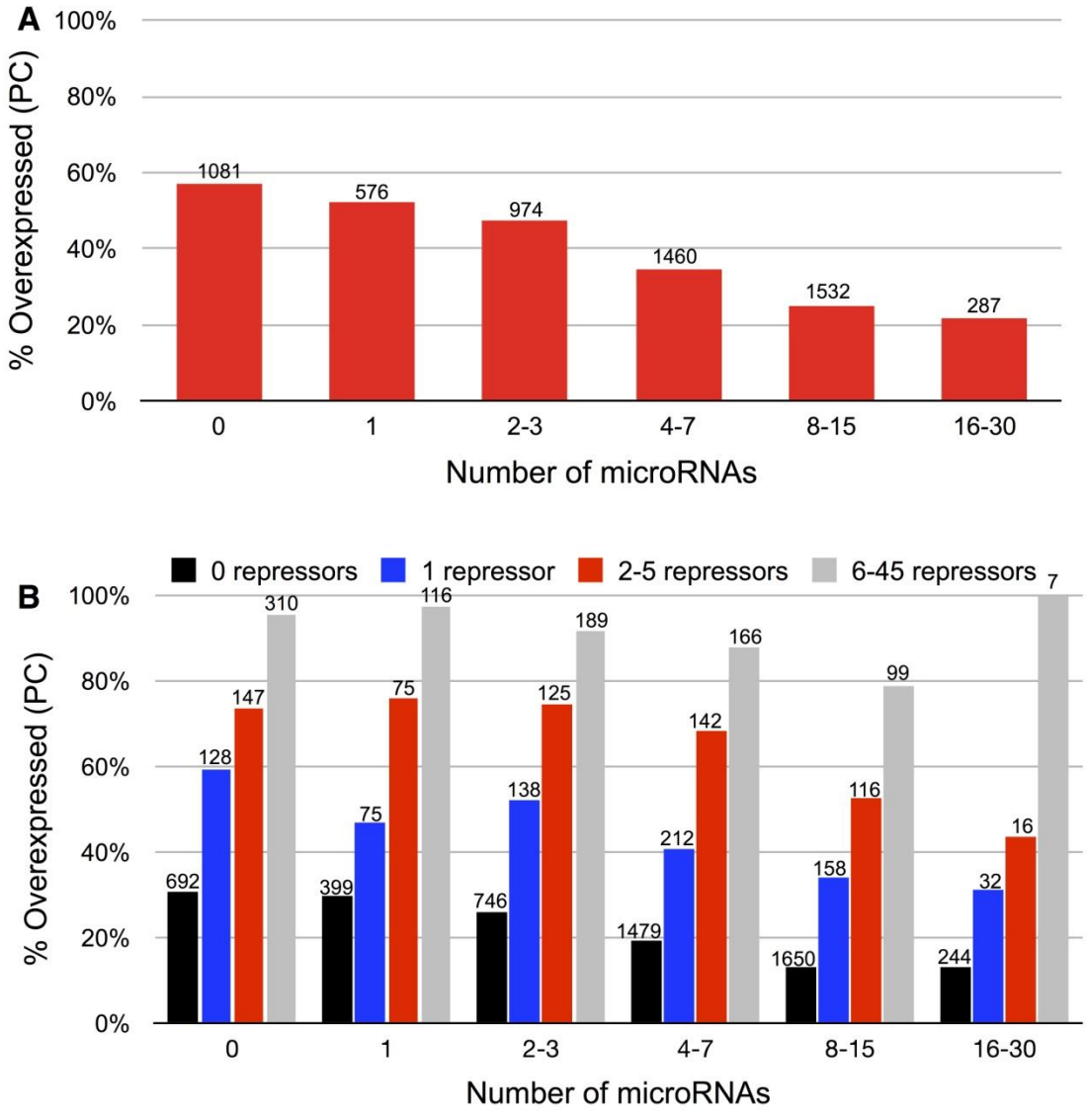


Figure 3.4: Analysis of global changes in gene expression is consistent with TOM. Differentially expressed genes (5910) targeted by upregulated miRNAs were divided into 6 groups based on the number of miRNAs targeting each gene (The number of genes included in each group are presented on top of the bars). A) An inverse relationship exists

between the number of upregulated miRNAs targeting genes and the percentage of these genes displaying the unexpected PC change in expression. The chi-square test for trend is ($X^2=311.5$; $p << .0001$), indicating a significant increase in down-regulated targets as the number of targeting microRNAs increases. B) The strength of miRNA regulatory control is diminished as the number of deregulated repressor genes increase (black bars = genes targeted by 0 repressors; blue bars = genes targeted by 1 repressor; red bars = genes targeted by 2-5 repressors; gray bars = genes targeted by 6-30 repressors). A chi-square test for trend for the miRNAs in this figure (3.4B) is ($X^2=444.6$; $p << .0001$), indicating a significant increase in down-regulated miRNA targets as the number of miRNAs increases. The chi-square test for trend for the transcriptional repressors in this figure (3.4B) is ($X^2=1904.6$; $p << .0001$), indicating a significant increase in upregulated targets as the number of repressors increases.

One approach taken by systems biologists to model regulatory relationships in complex cellular contexts is to use highly correlated changes in expression patterns among genes as evidence of direct and/or indirect interactions (Allocco, Kohane et al. 2004; (Mansson, Tsapogas et al. 2004). In our case, we examined variation in gene expression patterns across our OSE samples to identify genes displaying consistent inverse correlations (Pearson's $r < -0.8$) in expression with changes in expression of the 105 repressor genes previously characterized as significantly downregulated in EOC and regulatory targets of one-or-more of the 31 miRNAs (see above). Genes displaying an inversely correlated pattern of co-expression (n=1205) were operationally classified as targets of these repressor genes. Genes not displaying this pattern of expression (n=3624) were classified as non-targets of the designated repressor genes. Having established these classes, it became possible to distinguish between regulatory interactions fulfilling the triangular relationship of TOM from those that do not.

The results presented in Figure 3.4B again indicate that as the number of upregulated miRNAs targeting genes increases, the percentage of the unexpected PC decreases significantly (chi-square test for trend X^2=444.6, p<<.0001). The overlap of miRNA targets and down-regulated genes, as seen in the Venn diagram (Figure 3.5B), is also highly significant (hypergeometric p < 1E-12). However, consistent with TOM, the magnitude of this effect is significantly reduced as the number of downregulated repressor genes exerting an opposing effect increases (Figure 3.4A; chi-square test for trend X^2=1904.6, p<<.0001). The overlap of repressor targets and upregulated genes (Figure 3.5A) is also highly significant (hypergeometric p < 1E-12). Indeed, the results indicate that the derepression of target genes mediated by high (6-45) numbers of downregulated repressors is sufficient to nearly or completely override the regulatory controls of even large numbers (>15) of upregulated miRNAs.

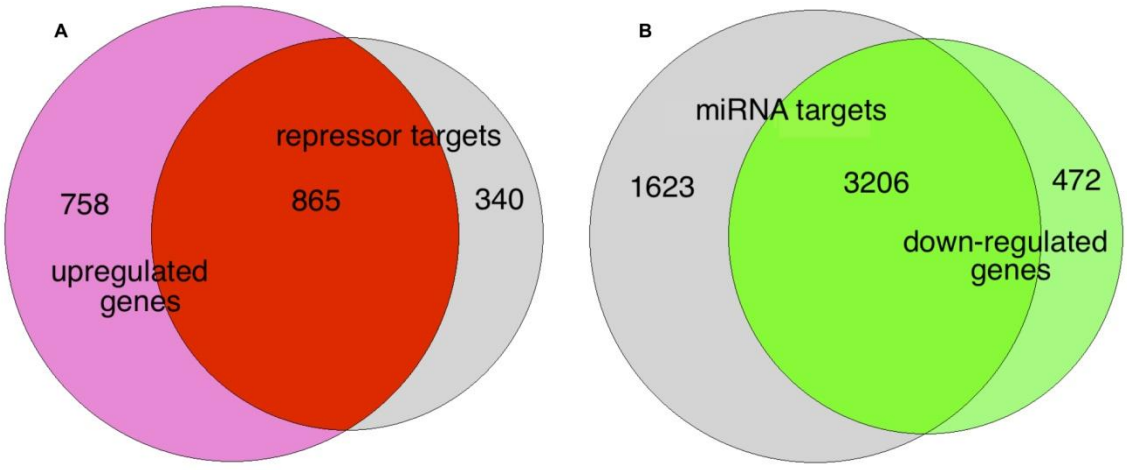


Figure 3.5: Upregulated repressor targets and down-regulated miRNA targets show significant enrichment in support of TOM. Venn diagrams show the enrichment of A) downregulated targets of miRNAs (Figure 3.4B), which was consistent with our transcriptional override model (TOM) using a hypergeometric distribution (p<1E-12, population=5910, total down-regulated genes =3678, number of predicted miRNA targets=4829, down-regulated miRNA targets=3206); and B) upregulated targets of

repressors (Figure 3.4B), which targets are also predicted targets of 1 or more upregulated miRNAs, and which we found consistent with TOM using a hypergeometric distribution ($p < 1E\text{-}12$, population=4829, total upregulated genes=1623, number of repressor targets=1205, upregulated repressor targets=865).

Discussion

Recent studies have clearly established miRNAs as early indicators of disease (Tijsen, Pinto et al. 2012; (Di Leva and Croce 2013) and as a potential new class of therapeutic agents (Shahab, Matyunina et al. 2012; (Gurtan and Sharp 2013). Full appreciation of the biological significance of modulations in levels of miRNAs, as well as the future rational employment of miRNAs as therapeutic agents, will require an understanding of both the direct and indirect molecular consequences of changes in the levels of miRNAs on cell function. While the direct gene (mRNA) targets of individual miRNAs can be computationally predicted and experimentally validated with varying degrees of accuracy (Ritchie, Rasko et al. 2013), reliable predictions of the indirect molecular effects of changes in miRNA levels have remained a major challenge in molecular systems biology (Gurtan and Sharp 2013; (Vera, Lai et al. 2013).

Testing the model's ability to globally predict the relative influence of miRNA and repressor gene regulatory controls on target gene expression is problematic for two reasons: first, a compendium of all human repressor genes and their regulatory targets is currently unavailable; second, many regulatory proteins can function as repressors or activators depending on cellular context and protein complex association (Remenyi, Scholer et al. 2004).

With that caveat, we present a regulatory network model (TOM) that explains a significant component of the unexpected low frequency of IC changes in expression levels between mRNAs and their regulating miRNAs. The model postulates that the expected down-regulation of target genes induced by elevated levels of regulating miRNAs may be masked or "overridden" by increases in transcriptional initiation mediated by the down-regulation of repressor genes that are themselves targets of the same regulating miRNAs (Figure 3.1A). Depending upon the strength of the transcriptional override (*i.e.*, the relative strengths of miRNA and repressor gene mediated derepression), TOM suggests that increases in miRNA levels may display no effect (NC) or be positively correlated (PC) with changes in levels of their targeted mRNAs. We show that the model successfully accounts for a majority of the unexpected relationships observed between changes in miRNA levels and levels of their targeted mRNAs measured in the same cohort of cancer epithelial cells isolated from a series of ovarian cancer patients (Additional file B.1).

While we have evaluated TOM within the context of its ability to account for global patterns of changes in gene expression, the model also provides a framework for explaining specific interactions between miRNAs and target genes (*e.g.*, Figure 3.1B). Thus, TOM may be of use in not only establishing the likely consequence of exogenous expression of miRNAs on specific target genes, but in predicting likely indirect global effects as well.

Conclusions

It is now widely acknowledged that the complexity of molecular interactions taking place on the cellular level can significantly obscure the expected consequences of molecular processes characterized *in vitro*, (Hornberg, Bruggeman et al. 2006; (Vidal, Cusick et al. 2011). Our findings indicate that the direct and indirect regulatory effects of changes in miRNA expression levels *in vivo* are interactive and complex but amenable to systems level modeling. We have shown that TOM can account for a major component of the unexpected consequences of changes in miRNA expression levels on their target mRNAs. Although the model has been developed and evaluated within the context of ovarian cancer, we believe it may be applicable in other biological contexts as well including the potential use in the rational design of miRNA-based strategies for the treatment of cancer and other diseases.

Methods

All tissues were collected according to previously published procedures (Bowen, Walker et al. 2009) following approved Institutional Review Board protocols from Northside Hospital (Atlanta) and Georgia Institute of Technology. Informed consent was obtained from all subjects. The histopathology for all cancer patients was serous papillary adenocarcinoma of the ovary and for the control patients the ovaries were considered within normal limits.

mRNA microarray data analysis

Ten OSE (normal) and ten EOC (cancer) samples were analyzed for mRNA expression using the Affymetrix Gene Chip Operating System (GCOS HG-U133 Plus 2.0). CEL

files generated by GCOS were converted to expression values using GCRMA normalization on the arrayanalysis.org (Eijssen, Jaillard et al. 2013) website, whose output also included quality control metrics, principal components analysis (PCA) and cluster dendrograms. Present/absent calls were generated from the MAS 5.0 statistical algorithm as implemented in Affymetrix Expression Console. Probe sets with >60% present calls in either of the two groups (OSE and EOC) were selected for further analysis. After log2 transformation, signal values of those probe sets were submitted to Statistical Analysis of Microarrays (SAM) for multiple testing correction where a 5.5% FDR was applied resulting in 7462 probe sets representing 5910 differentially expressed genes (DEGs). Annotations for probe sets were obtained from Affymetrix (Affymetrix). The processed and raw data files for the samples used in this study have been deposited in the Gene Expression Omnibus (GSE52037 with superseries GSE52460).

microRNA microarray data analysis

Expression profiles for miRNAs from three OSE and three EOC samples were generated by Asuragen (Austin, TX) using Ambion miRChip technology (Life Technologies). Two sets of CEL files, created from 6 biological replicates and two sets of technical replicates were normalized using MAS 5.0 to expression signals, giving 6 values per probe/gene. Probe sets labeled as human (those having an "hsa-" prefix), known to be conserved to mouse (mmu), and with at least 65% present calls (calculated by Asuragen) in either of the two groups (OSE and EOC) were selected for further filtering. Thirty-one differentially expressed miRNAs (fc > 6, p-value < .03) were selected. The repressive potential of all 31 miRNAs was validated by noting that > 65% of the predicted DEG

targets of each upregulated miRNA were actually down-regulated, while only 44% of DEGs *not* predicted to be targets of any upregulated miRNA were downregulated. Mean repression over all 31 miRNAs was 71%. The processed and raw data files for the samples used in this study have been deposited in the Gene Expression Omnibus (GSE52459 with superseries GSE52460).

microRNA target prediction

The miRNA target prediction file based on mirSVR was downloaded from miRNA.org (August 2010 release). The mirSVR score refers to targets of miRNAs with scores obtained from their support vector regression algorithm. To reduce the occurrence of false positives, only predicted targets with a mirSVR score less than -.2 were considered. The miRNA target predictions based on TargetScan and SVMicrO were downloaded from www.targetscan.org (retrieved 8/2010) and www.compgenomics.utsa.edu/Result/Human/hsa_human (retrieved 9/2010), respectively.

Transcriptional repressor selection

Members of the Gene Ontology categories GO:0045892, GO:0000122, GO:0010944, GO:0032088 and GO:0008156, relating to the negative-regulation-of-transcription or its child terms, were downloaded from the European Bioinformatics Institute (EBI) and parsed using UNIX scripts. In that download, we found 439 potential repressor genes. Of those, 109 genes were significantly down-regulated according to our microarray analysis and 105 of these genes were also predicted targets of one or more of the 31 upregulated

miRNAs (v.s.). These 105 transcriptional repressor genes formed the basis for miRNA target derepression in our model.

Transcriptional repressor target prediction and experimental validation

To obtain predicted and/or experimentally validated transcription factor binding site data, we downloaded the TRANSFAC data file **c3.tft.v3.1.symbols.gmt** from GSEA (Gene Set Enrichment Analysis website - http://www.broadinstitute.org/gsea/downloads.jsp). Data files were parsed with UNIX scripts, which extracted pairs of genes consisting of one repressor and one or more binding partners. All repressor-partner pairs under consideration had to be DEGs and predicted targets of at least one of the 31 upregulated miRNAs, and all transcriptional repressors were down-regulated in cancer. Further, all repressor-partner pairs were required to show a correlation coefficient of $r < -.8$ across all normal samples.

Correlation coefficient calculation

For the global analysis of relationships among all 105 transcriptional repressors and their binding partners, Pearson's correlation coefficient (PCC) was calculated across all ten OSE (normal) samples between all transcriptional repressors and predicted miRNA targets. Specifically, we used the Mathematica (Wolfram Research 2010) correlation[] function ($n=10$; $r < -.8$) for a directional significance of ($p < .0027$). Fold-change from normal to cancer in these genes ranged from -625 to 121.

Availability of supporting data

The processed and raw data files for the samples used in the mRNA and miRNA expression studies have been deposited in the Gene Expression Omnibus (GSE52037 and GSE52459 with Super-series GSE52460).

CHAPTER 4

EVIDENCE OF SIGNIFICANT CHANGES IN GENE NETWORK INTERACTIONS IN OVARIAN CANCER

Introduction

It has become increasingly clear in recent years that the molecular basis of cancer and other complex human diseases is not merely attributable to structural and/or regulatory mutations in one or a few genes but additionally to disruptions in networks of regulatory interactions existing among these and other components of the genome (Ideker, Galitski et al. 2001; (Vidal, Cusick et al. 2011; (Wu, Zhu et al. 2012; (Ying, Lv et al. 2013). Although an experimental determination of the myriad of potential genetic interactions that characterize human cells is only beginning to be established experimentally e.g., (Vidal, Cusick et al. 2011), a number of computational approaches have recently been developed that can directly or indirectly infer interactive network relationships among genes in normal and cancer cells (Laubenbacher and Stigler 2004; (Schafer and Strimmer 2005; (Kar, Gursoy et al. 2009; (Barabasi, Gulbahce et al. 2011). The standard input for gene regulatory network algorithms is gene expression (RNA-seq and/or microarray) datasets. Interactions (edges) between genes (nodes) can be inferred in an unsupervised fashion based upon consistent and highly correlated changes (positive or negative) in the expression patterns that exist among genes in multiple biological samples (Roth, Hevezi et al. 2006; (Amar, Safer et al. 2013). In this study, we employ an unsupervised (Pearson correlation) approach to estimate overall gene network relationships between precursor

(control) ovarian surface epithelial cells (OSE) and serous papillary ovarian cancer epithelial cells (CEPI) isolated from human patient samples by laser capture microdissection (LCM) and analyzed by gene expression microarray (Affymetrix, U133 Plus 2) as previously described (Bowen, Walker et al. 2009). The initial results indicate that ovarian cancers display a significant overall reduction in correlated gene interactions relative to normal precursor cells. Consistent with this finding, we further show that biological processes identified by gene ontology (GO) analyses to be significantly over-represented among differentially expressed genes were substantially different from those overrepresented among genes involved in qualitatively disrupted network interactions. Although many of the genes associated with qualitatively disrupted network interactions have been previously implicated in ovarian and other cancers, others may represent potential new targets for therapeutic intervention.

Results

Ovarian cancer is associated with a significant reduction in gene network interactions relative to normal precursor cells.

We employed microarray gene expression analysis to compare differences in gene expression levels between precursor ovarian surface epithelial (OSE) cells and ovarian cancer epithelial cells (CEPI) isolated from patient tissue samples by laser capture microdissection (LCM). Gene expression profiling identified 5,070 significantly differentially expressed genes (mRNAs) between the OSE and CEPI samples (Table C.1). Of these, 1,436 (28%) were significantly up-regulated and 3,634 (72%) significantly down-regulated in the cancer samples relative to normal controls.

Baseline relationships (Pearson correlations) between significantly differentiated genes in the control samples (OSE) were established by comparing the expression levels of 12.8 million pairs of genes (5,070 X 5,069 / 2) across 10 normal patient samples and selecting gene pairs displaying highly correlated (positive or negative) expression levels across samples. This resulted in a total of 682,391 (< 5% of all possible pairings) significantly correlated (positive and negative) pairs of differentially expressed genes ($r > 0.85$; $p < .0018$; Note: we determined that (r) < 0.85 can occur purely by chance - Figure 4.1).

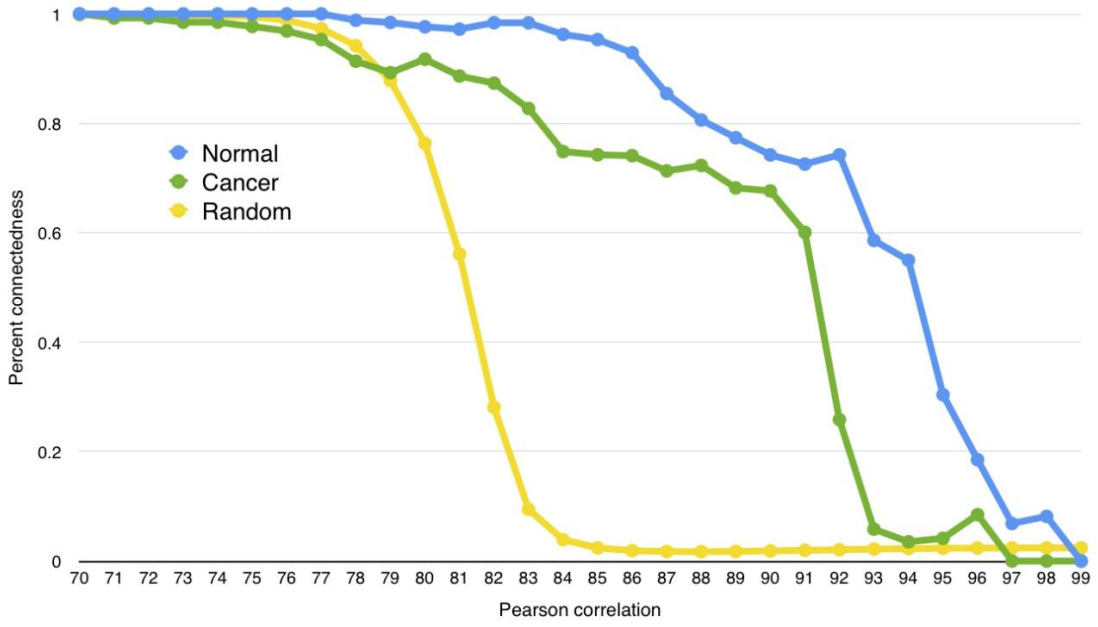


Figure 4.1: Comparison of normal, cancer and random network connectivity. Each point on the three lines of this graph represents the size of the largest connected sub-network as a percentage of the total number of correlated genes, across 10 samples, at that r value (.70 < r < .99). The graph compares connectedness among correlated DE genes using normal (dark blue line), cancer (green line) and 500 randomly generated (yellow) expression values across different values of r. This comparison shows that even random expression values can generate apparently highly correlated networks at r < .85. Based on this analysis, r was set to .85 to minimize the possibility of unjustifiable correlations.

Of these 682,391 correlated gene pairs, 568,083 were positively correlated with one another while 114,505 were negatively correlated. Conducting the same analysis across the 10 cancer (CEPI) samples resulted in less than 193,288 highly correlated (positive and negative) gene pairs ($r > 0.85$, $p<0.0018$) implying a substantial net loss in regulatory control in the ovarian cancer samples. Of the nearly 193,000 highly correlated gene pairs in cancer, 164,295 were positively correlated with one another while 29,005 were negatively correlated. Thus, only ≈35,000 pairs of genes (~5% of 682,391) were found to maintain the same patterns of expression between one another in the normal and cancer samples. These findings are consistent with previous experimental evidence indicating that cancer cells are generally associated with a significant loss in regulatory control relative to normal precursor cells; e.g., (Ford and Pardee 1999; (Zhao, Sun et al. 2012; (Cordero, Sole et al. 2014).

More than 45,000 gene pair relationships identified in normal samples qualitatively change in cancer samples.

While the above results indicate that a majority of gene pair interactions are disrupted in cancer, the findings are difficult to functionally interpret because relatively small quantitative changes in the expression levels of gene pairs between normal and cancer samples can significantly affect the gain or loss of pair-wise correlations, many of which may not be of biological significance. For this reason, we limited our subsequent analyses to gene pairs displaying qualitative reversals in interactive relationships (positive to negative or negative to positive) between the normal and cancer samples. Such qualitative reversals in regulatory controls have been previously shown to be associated

with functionally significant changes in cancer; e.g., (Oyake, Itoh et al. 1996; (Remenyi, Scholer et al. 2004; (Gordon, Akopyan et al. 2006; (Thompson, Xu et al. 2009). Gene pairs displaying the same qualitative patterns in normal and cancer samples, i.e., gene pairs with expression patterns that are consistently positively or negatively correlated in both normal and cancer samples, are classified as "consistent". Conversely, gene pairs with expression patterns that display qualitatively different relationships in the normal and cancer samples, i.e., gene pairs that display positively correlated relationships in normal tissues but negatively correlated relationships in cancer tissues and *vice* versa, are classified as "inconsistent". Of the 568,083 positively correlated gene pairs detected in the normal samples, 534,011 (94%) were found to maintain this same qualitative relationship in the cancer samples ("consistent"). Similarly, of the 114,505 negatively correlated gene pairs detected in normal samples, 103,277 (90%) were found to maintain this qualitative relationship in cancer (consistent). Thus, a total of only 45,300 (6.6%) gene pairs displayed a reversal in gene pair relationships in cancer ("inconsistent").

Genes associated with inconsistent gene pair relationships in cancer are significantly enriched for cell cycle and cell death related functions.

To determine if genes associated with inconsistent network relationships in our cancer samples were enriched for specific biological functions, we selected from the 45,300 pairs of genes involved in at least one inconsistent interaction, 350 genes associated with the greatest number of inconsistent interactions (Table C.2) and submitted them to DAVID (**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery) (Huang da, Sherman et al. 2009) for pathway enrichment analysis . The results (Table 4.1) identified

cell cycle related functions (red) as the most significantly enriched biological processes in this group of genes followed closely by functions associated with programmed cell death (green) (Figure 4.1a,b).

Table 4.1: Gene ontology enrichment of the top 350 inconsistent genes. The 350 genes involved in the greatest number of inconsistent pair-wise interactions across all highly correlated genes were analyzed for biological process enrichment using DAVID. M Phase ($p < 9.38E-10$) and other biological processes related to cell cycle (green) were found to be the most significant processes. Apoptosis and 11 other related processes (red), including Regulation of Programmed Cell Death were also highly enriched ($p < 0.05$).

Term	Gene Count	%	P-value
GO:0000279~M_phase	28	8.12	9.38E-10
GO:0000280~nuclear_division	21	6.09	3.15E-08
GO:0007067~mitosis	21	6.09	3.15E-08
GO:0022403~cell_cycle_phase	29	8.41	3.50E-08
GO:0000087~M_phase_of_mitotic_cell_cycle	21	6.09	4.28E-08
GO:0048285~organelle_fission	21	6.09	6.20E-08
GO:0007049~cell_cycle	40	11.59	1.79E-07
GO:0022402~cell_cycle_process	32	9.28	6.83E-07
GO:0000278~mitotic_cell_cycle	25	7.25	7.66E-07
GO:0051301~cell_division	21	6.09	3.26E-06
GO:0051276~chromosome_organization	24	6.96	1.26E-04
GO:0007059~chromosome_segregation	9	2.61	2.64E-04
GO:0000070~mitotic_sister_chromatid_segregation	6	1.74	8.63E-04
GO:0000819~sister_chromatid_segregation	6	1.74	9.80E-04
GO:0051726~regulation_of_cell_cycle	17	4.93	1.42E-03
GO:0033554~cellular_response_to_stress	24	6.96	1.51E-03
GO:0006259~DNA_metabolic_process	22	6.38	1.70E-03
GO:0006974~response_to_DNA_damage_stimulus	18	5.22	1.91E-03
GO:0007093~mitotic_cell_cycle_checkpoint	6	1.74	1.96E-03
GO:0000075~cell_cycle_checkpoint	8	2.32	2.46E-03
GO:0007010~cytoskeleton_organization	19	5.51	3.71E-03
GO:0010608~posttranscriptional_regulation_of_gene_expression	12	3.48	4.45E-03
GO:0007346~regulation_of_mitotic_cell_cycle	10	2.90	4.53E-03
GO:0051130~positive_regulation_of_cellular_component_organization	11	3.19	4.67E-03
GO:0010604~positive_regulation_of_macromolecule_metabolic_process	30	8.70	5.78E-03
GO:0006417~regulation_of_translation	9	2.61	7.21E-03
GO:0060541~respiratory_system_development	8	2.32	7.23E-03
GO:0090003~regulation_of_establishment_of_protein_localization_to_plasma_membrane	3	0.87	8.57E-03
GO:0010564~regulation_of_cell_cycle_process	8	2.32	9.63E-03
GO:0030071~regulation_of_mitotic_metaphase/anaphase_transition	4	1.16	1.04E-02
GO:0006916~anti-apoptosis	11	3.19	1.06E-02
GO:0007051~spindle_organization	5	1.45	1.43E-02
GO:0006281~DNA_repair	13	3.77	1.44E-02
GO:0031328~positive_regulation_of_cellular_biosynthetic_process	24	6.96	1.45E-02
GO:0001568~blood_vessel_development	12	3.48	1.47E-02
GO:0006260~DNA_replication	10	2.90	1.50E-02
GO:0006917~induction_of_apoptosis	14	4.06	1.59E-02
GO:0010557~positive_regulation_of_macromolecule_biosynthetic_process	23	6.67	1.61E-02
GO:0012502~induction_of_programmed_cell_death	14	4.06	1.63E-02
GO:0009891~positive_regulation_of_biosynthetic_process	24	6.96	1.69E-02
GO:0030324~lung_development	7	2.03	1.70E-02
GO:0001944~vasculature_development	12	3.48	1.73E-02
GO:0043068~positive_regulation_of_programmed_cell_death	17	4.93	1.83E-02
GO:0006310~DNA_recombination	7	2.03	1.86E-02
GO:0010942~positive_regulation_of_cell_death	17	4.93	1.92E-02
GO:0030323~respiratory_tube_development	7	2.03	1.94E-02
GO:0042770~DNA_damage_response_signal_transduction	6	1.74	2.50E-02
GO:0035196~gene_silencing_by_miRNA_production_of_miRNAs	3	0.87	2.51E-02
GO:0051783~regulation_of_nuclear_division	5	1.45	2.79E-02
GO:0007088~regulation_of_mitosis	5	1.45	2.79E-02
GO:0031050~dsRNA_fragmentation	3	0.87	2.93E-02
GO:0043067~regulation_of_programmed_cell_death	26	7.54	3.02E-02
GO:0006333~chromatin_assembly_or_disassembly	7	2.03	3.04E-02
GO:0046324~regulation_of_glucose_import	4	1.16	3.13E-02
GO:0010941~regulation_of_cell_death	26	7.54	3.20E-02
GO:0010827~regulation_of_glucose_transport	4	1.16	3.38E-02
GO:0043065~positive_regulation_of_apoptosis	16	4.64	3.42E-02
GO:0033043~regulation_of_organelle_organization	10	2.90	3.47E-02
GO:0048514~blood_vessel_morphogenesis	10	2.90	3.47E-02
GO:0001817~regulation_of_cytokine_production	9	2.61	3.60E-02
GO:0035088~establishment_or_maintenance_of_apical/basal_cell_polarity	3	0.87	3.84E-02
GO:0007052~mitotic_spindle_organization	3	0.87	3.84E-02
GO:0042116~macrophage_activation	3	0.87	3.84E-02
GO:0006302~double-strand_break_repair	5	1.45	4.08E-02
GO:0043137~DNA_replication_removal_of_RNA_primer	2	0.58	4.15E-02
GO:0033567~DNA_replication_Okazaki_fragment_processing	2	0.58	4.15E-02
GO:0006915~apoptosis	20	5.80	4.17E-02
GO:0042552~myelination	4	1.16	4.19E-02
GO:0051271~negative_regulation_of_cell_motion	5	1.45	4.29E-02
GO:0045765~regulation_of_angiogenesis	5	1.45	4.29E-02
GO:0051252~regulation_of_RNA_metabolic_process	47	13.62	4.50E-02
GO:0006357~regulation_of_transcription_from_RNA_polymerase_II_promoter	23	6.67	4.67E-02
GO:0012501~programmed_cell_death	20	5.80	4.70E-02

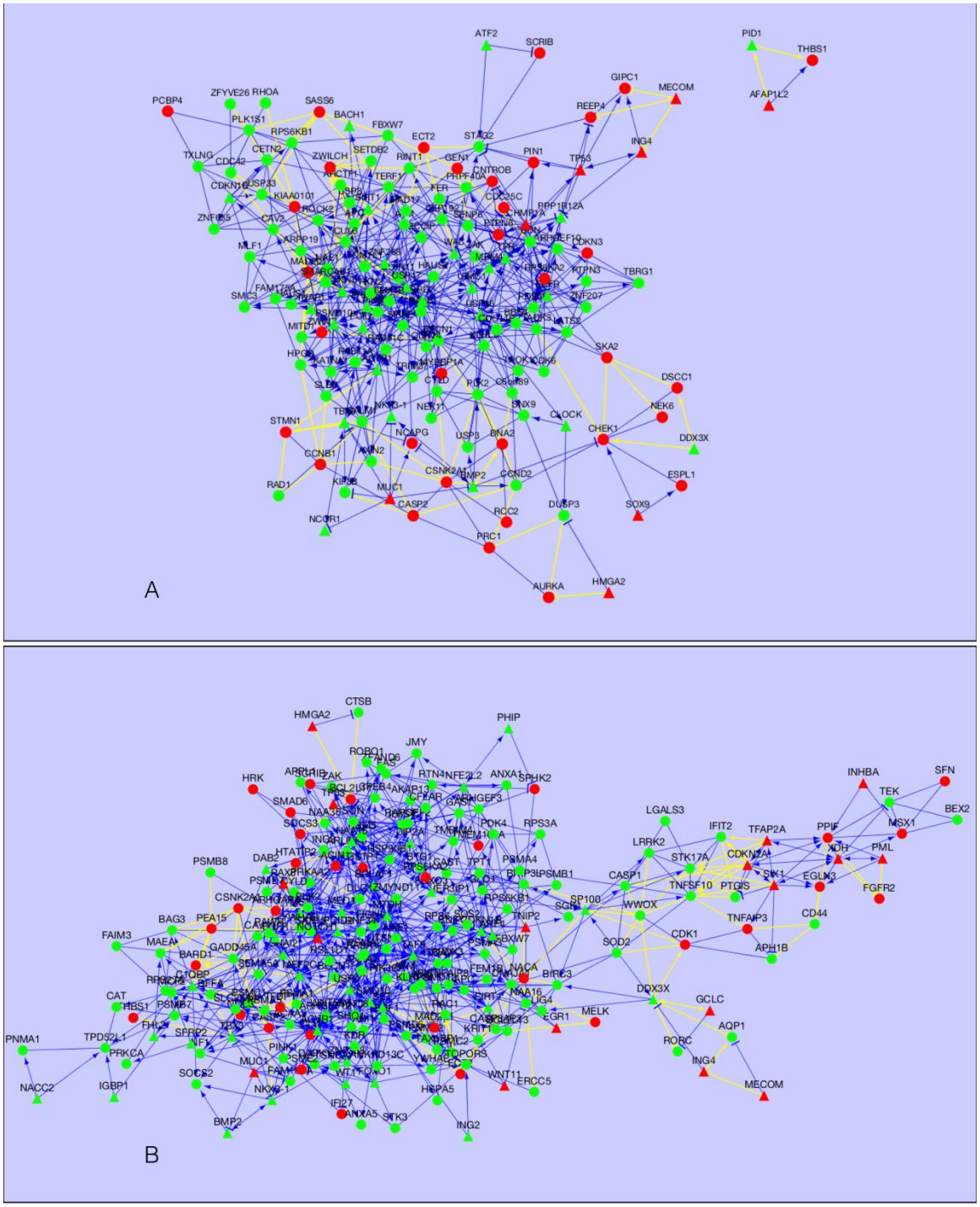


Figure 4.2: Networks of enriched genes. Two networks showing consistent (blue) and inconsistent (yellow) interactions (edges) between DE genes (nodes) as those interactions appear in cancer. Red nodes represent genes upregulated in cancer, while green nodes are down-regulated. Genes were selected based on their membership in either (A) cell cycle (GO:0007049) or (B) regulation of programmed cell death (GO:0045076). The (A) network shows 598 pairs of cell cycle genes which were highly correlated ($r > 0.85$)

across 10 normal samples, while the (B) network shows 924 pairs of highly correlated genes related to programmed cell death.

Genes associated with qualitatively disrupted network interactions in ovarian cancer identify a potential new class of candidates for targeted gene therapy.

A current molecular standard used to identify functionally significant genetic changes in cancer is changes in gene expression (Welsh, Sapinoso et al. 2001). To determine the extent to which processes identified as significantly overrepresented among genes involved in disrupted network interactions (see above) were similar or different from processes overrepresented in genes most significantly differentially expressed, we selected the 350 most significantly differentially expressed genes between the normal and cancer samples (Table C.3) and subjected them to the gene ontology analysis. As shown in Table 4.2, the biological processes most significantly overrepresented among differentially expressed genes was substantially different from those most significantly overrepresented among differentially expressed genes involved in disrupted network interactions.

Table 4.2: Gene ontology enrichment of the top 350 differentially expressed genes. The 350 genes with the greatest fold change from normal to cancer (FC < -9.21 and 9.27 < FC) were analyzed for biological process enrichment using DAVID. Cell adhesion ($p < 9.25E-7$) was found to be the most significantly enriched process.

Term	Gene Count	%	P-value
GO:0007155~cell_adhesion	36	10.50	9.25E-07
GO:0022610~biological_adhesion	36	10.50	9.60E-07
GO:0048732~gland_development	13	3.79	2.44E-05
GO:0060429~epithelium_development	15	4.37	1.22E-04
GO:0006928~cell_motion	24	7.00	1.24E-04
GO:0043627~response_to_estrogen_stimulus	10	2.92	3.49E-04
GO:0030182~neuron_differentiation	21	6.12	7.17E-04
GO:0030855~epithelial_cell_differentiation	10	2.92	8.26E-04
GO:0010033~response_to_organic_substance	29	8.45	1.19E-03
GO:0007411~axon_guidance	9	2.62	1.41E-03
GO:0048812~neuron_projection_morphogenesis	13	3.79	1.44E-03
GO:0032989~cellular_component_morphogenesis	19	5.54	1.47E-03
GO:0009611~response_to_wounding	23	6.71	1.49E-03
GO:0009725~response_to_hormone_stimulus	18	5.25	2.02E-03
GO:0001655~urogenital_system_development	9	2.62	2.28E-03
GO:0009719~response_to_endogenous_stimulus	19	5.54	2.35E-03
GO:0048545~response_to_steroid_hormone_stimulus	12	3.50	2.45E-03
GO:0031175~neuron_projection_development	14	4.08	2.46E-03
GO:0050878~regulation_of_body_fluid_levels	10	2.92	2.75E-03
GO:0030204~chondroitin_sulfate_metabolic_process	4	1.17	2.85E-03

Indeed, of the 350 genes most significantly differentially expressed between our normal and cancer samples, only 22 overlapped with the 350 genes involved in the most frequently disrupted pair-wise interactions (Figure 4.1, Table C.4).

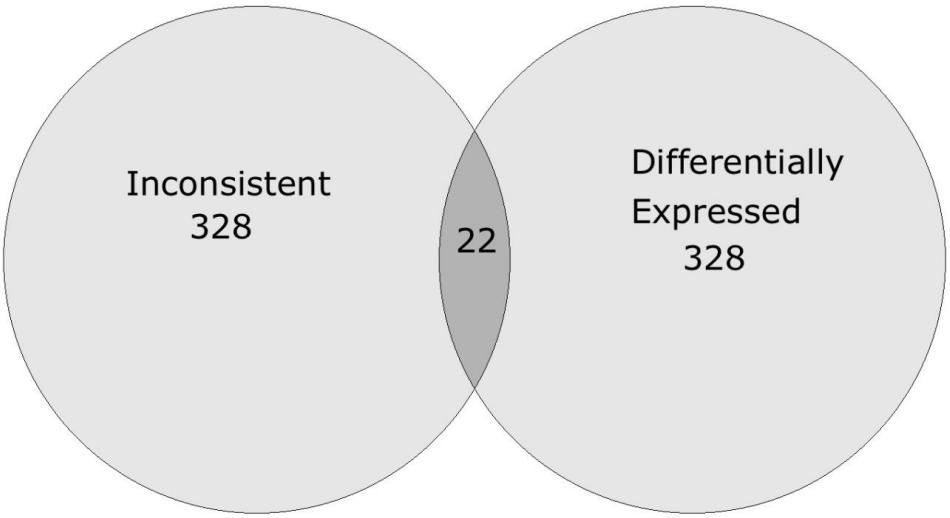


Figure 4.3: Venn diagram showing overlap between differentially expressed vs. inconsistent genes. The light gray circle on the left represents 350 genes which were involved in the greatest number of inconsistent pairs in normal samples, while the circle on the right represents 350 most differentially expressed genes. Only 22 genes are common to both groups, which, by hypergeometric distribution ($p < .687$) is not significant.

These results suggest that the identification and characterization of genes involved in disrupted gene network interactions in ovarian and perhaps other cancers may constitute an important source of functionally significant information unavailable through gene expression analyses alone.

Many of the genes associated with inconsistent interactions have been previously implicated in cancer.

We found that many of the genes involved in inconsistent interactions in ovarian cancer have been previously associated with the cancer (Table 4.3).

Table 4.3: Inconsistent genes implicated in cancer.
This table shows that of 34 genes involved in large numbers of inconsistent pairings, 29 have been previously implicated in ovarian cancer and 5 others have been reported in other cancers. Column 1 indicates whether the gene is a known drug target, according to the CancerResource (Ahmed, Meinel et al. 2011). Columns 2-6 show the gene symbol, fold change, number of inconsistent interactions and reference, respectively.

Druggable	Gene symbol	Fold change	Inconsistent interactions	Cancer	Reference
Y	AURKA	16.9	51	ovarian	(Wrzeszczynski, Varadan et al. 2011)
Y	BARD1	2.1	103	ovarian	(Deng 2006)
Y	CCNA2	2.9	221	ovarian	(Bowen, Walker et al. 2009)
Y	CCNB1	5.7	210	ovarian	(Bast 2011)
Y	CDK1	9.6	79	ovarian	(See, Kavanagh et al. 2003)
Y	CHEK1	4.0	187	ovarian	(Bertoni, Codegoni et al. 1999)
Y	COL1A2	2.8	612	ovarian	(Karlan, Dering et al. 2014)
N	COL5A2	2.1	419	ovarian	(Helleman, Jansen et al. 2006)
N	DNA2	1.7	209	ovarian	(Miles, Seiler et al. 2012)
N	EZR	-41.8	398	ovarian	Poersch '14
N	FANCD2	7.0	209	ovarian	(Lewis, Flanagan et al. 2005)
N	FEN1	2.4	56	ovarian	(Abdel-Fatah, Russell et al. 2014)
N	FGFR1OP	2.2	322	ovarian	(Huang, Chen et al. 2012)
N	HMGB3	3.7	430	ovarian	(Helleman, Jansen et al. 2006)
Y	HTATIP2	2.1	128	ovarian	(Kumtepe, Halici et al. 2013)
Y	KIAA0101	12.7	118	ovarian	(Emmanuel, Gava et al. 2011)
Y	MAD2L1	2.8	263	ovarian	(Park, Jeong et al. 2013)
N	MDC1	1.6	54	ovarian	(Deng 2006)
Y	MECOM	18.6	20	ovarian	(Ramakrishna, Williams et al. 2010)
Y	PRC1	3.7	56	ovarian	(Ehrlichova, Mohelnikova-Duchonova et al. 2013)
N	PTPRM	4.2	506	ovarian	(Gyorffy, Dietel et al. 2008)
Y	RAD51AP1	8.1	88	ovarian	(Miles, Seiler et al. 2012)
N	RAD52	1.8	169	ovarian	(Auranen, Song et al. 2005)
N	RMI1	1.7	495	ovarian	(Moolthiya and Yuenyao 2009)
Y	STMN1	1.6	177	ovarian	(Ying, Su et al. 2013)

Table 4.3 (continued)

N	TIMELESS	3.0	61	ovarian	(Tokunaga, Takebayashi et al. 2008)
Y	TOPBP1	1.8	116	ovarian	(Karppinen, Erkko et al. 2006)
Y	WWOX	-2.0	189	ovarian	(Gourley, Paige et al. 2005)
Y	XPR1	2.3	645	ovarian	(Nikolova, Doganov et al. 2009)
N	CSNK2A1	1.8	84	pancreatic	(Giroux, Iovanna et al. 2006)
Y	HELLS	2.2	576	NSCLC	(Yano, Ouchida et al. 2004)
N	TET1	1.6	574	prostate	(Hsu, Peng et al. 2012)
Y	VRK1	1.9	650	breast	(Valbuena, Castro-Obregon et al. 2011)
N	ZWILCH	1.6	294	breast	(Brendle, Brandt et al. 2009)

For example, CHEK1 (checkpoint kinase 1), a regulator of DNA damage-induced cell cycle (G2/M) arrest (Ouyang, Yao et al. 2009) and associated with 187 inconsistent interactions in our cancer samples (Table C.2), has been previously identified as regulator of tumor suppressor functions and as a potential target for cancer gene therapy (Helleday, Petermann et al. 2008; (Khanna, Kauko et al. 2013). MECOM (MDS1 And EVI1 Complex Locus) has been shown to be commonly amplified in ovarian cancer and has also been identified as a potential target for cancer gene therapy (Ramakrishna, Williams et al. 2010; (Cancer Genome Atlas Research 2011).

A number of other genes involved in inconsistent interactions in our ovarian cancer samples, although not yet implicated in ovarian cancer, have been identified as important contributors to other types of cancer. For example, the gene involved in the greatest number of inconsistent interaction in ovarian cancer (650), VRK1 (vaccinia-related kinase-1), has been previously associated with cell cycle and cell death functions (Valbuena, Castro-Obregon et al. 2011) and has recently been implicated in breast cancer

(Molitor and Traktman 2013). Likewise, the TET1 (*ten-eleven translocation 1 gene*) gene, which we found to be involved in 574 inconsistent interactions in OC, has been recently identified as a tumor suppressor gene associated with prostate cancer (Hsu, Peng et al. 2012). Another tumor suppressor gene, HELLS found to be involved in 576 inconsistent interactions in OC has been previously associated with epigenetic deregulation leading to lung cancer onset and/or progression (Yano, Ouchida et al. 2004). A number of the genes involved in inconsistent interactions in our ovarian cancer samples, although not yet explicitly implicated in any human cancer, are nevertheless known to be involved in important cellular functions and thus may represent a currently under appreciated class of cancer associated genes. For example, XPR1 (xenotropic and polytropic retrovirus receptor-1) is an atypical trans-membrane signaling receptor associated with G-protein coupled receptor activity (Vaughan, Mendoza et al. 2012). Although XPR1 is currently functionally recognized only as a receptor for xenotropic and polytropic retroviruses this is clearly not its primary cellular function. The fact that XPRI is associated with 645 inconsistent interactions in our ovarian cancer samples suggests that it may play a significant, albeit as-yet-unrecognized role in ovarian cancer onset and/or progression.

Discussion

Modern high-throughput sequencing and gene expression technologies are providing unprecedented insights into the molecular genetic processes disrupted in cancer, e.g., (Cancer Genome Atlas Research 2011; (Yates and Campbell 2012) and have led to the possibility of a personalized, targeted gene approach to cancer therapy (Dachs,

Dougherty et al. 1997). However, since not all variants detected by high-throughput technologies may be of functional significance, high-throughput sequencing and gene expression data must be coupled with the capability to distinguish functionally important or "driver" variants from "passenger" variants of little on no relevance. A major barrier to the attainment of this goal is the fact that the eukaryotic genome is a highly integrated network; e.g., (Barabasi and Oltvai 2004). As a consequence, the biological significance of a genetic variant in any particular cancer can reasonably be expected to depend, to a greater or lesser extent, upon the functional status of other genes with which it interacts. The growing realization of the potential importance of gene interactions in a variety of diseases; e.g., (Barabasi 2007), including cancer, e.g., (Laufer, Fischer et al. 2013; (Hill, Matyunina et al. 2014; (Kim, Li et al. 2014) is spurring the development of new and relevant experimental and computational tools to identify and evaluate the significance of changes in gene network relationships in eukaryotic cells; e.g., (Djebbari and Quackenbush 2008; (Laubenbacher and Jarrah 2009; (Nagaraj and Reverter 2011; (Moreau and Tranchevent 2012). Interactions between genes are typically computationally inferred in an unsupervised fashion based upon consistent and highly correlated changes (positive or negative) in the expression patterns between genes in multiple biological samples; e.g., (Soneson, Lilljebjorn et al. 2010; (Madhamshettiwar, Maetschke et al. 2012; (Stephens, Tarpey et al. 2012). Although, such unsupervised approaches typically infer network relationships exclusively from minimally processed gene expression data, more supervised methods (e.g., Bayesian (Kang, Zheng et al. 2011), partial correlation (Madhamshettiwar, Maetschke et al. 2012) etc.) incorporate additional supplementary information, such as current understanding of specific gene

regulatory relationships. While supervised methods can have advantages when inferring network changes in previously established, well-defined pathways, they can be less effective in detecting previously unrecognized interactions and/or in detecting important yet previously undefined system-wide changes in network structure. For example, supervised algorithms using Bayesian networks typically operate on directed acyclic graphs (DAG), that do not include loops from the network. Such procedures can result in the loss of biologically relevant data since feedback loops are widely recognized as biologically significant eukaryotic cells in general, e.g., (Pomerening 2009) and particularly in cancer; e.g., (Harris and Levine 2005; (Su, Liu et al. 2014) .

In this study, we were interested in exploring the possible significance of changes in gene interactions in ovarian cancer by comparing differences in correlated patterns of gene expression between precursor ovarian surface epithelial (OSE) cells and ovarian cancer epithelial cells (CEPI) isolated from 10 normal and 10 cancer patient samples. Using an unsupervised computational approach (Pearson correlation), we found evidence of a significant overall loss in gene-gene interactions in our cancer samples. These results are consistent with earlier computational and experimental studies suggesting that cancers are generally associated with an overall reduction in regulatory control (Ford and Pardee 1999; (Zhao, Sun et al. 2012; (Cordero, Sole et al. 2014).

In an effort to associate the most significant changes in gene pair interactions with specific biological functions, we developed a novel approach to identify gene pair interactions that qualitatively change in cancer samples. We found that genes involved in qualitative changes in gene pair interactions in our cancer samples were most significantly associated with biological processes involved in cell cycle regulation and

cell death. Interestingly, the genes identified as most frequently involved in these processes were not among the genes most significantly differentially expressed between our cancer and control samples. These results suggest that the identification of genes involved in disrupted network interactions in cancer may represent a clinically significant class of genes that may go undetected in standard mutational and gene expression analyses.

The long-term hope of converting cancer from a lethal to a manageable chronic disease rests not only upon the availability of technologies to accurately detect structural and expression genetic variants associated with individual tumors but analytical methods that can reliably identify those variants that are causally responsible for the cancer phenotype in individual patients. Our results indicate that the simple identification of genes mutated and/or differentially expressed in cancer tissues may not be sufficient to identify all genes of functional significance. The development and implementation of novel methods to identify genes involved in disrupted network interactions in cancer promises to provide important new insight into the processes underlying cancer and the potential identification of a new class of genes for targeted gene therapy.

Methods

Gene expression profile

We downloaded gene expression data from the Gene Expression Omnibus (GEO accession: GSE52037). The expression profiles were performed on ten normal ovarian surface epithelial (OCE) and ten serous papillary cancer epithelial (CEPI) samples (see (Hill, Matyunina et al. 2014) for details) and analyzed for changes in mRNA expression using the Affymetrix Gene Chip Operating System (GCOS HG-U133 Plus 2.0). CEL files generated by GCOS were converted to expression values using GCRMA normalization on the arrayanalysis.org [33] website, whose output also included quality control metrics and cluster dendrograms. Present/absent calls were generated from the MAS 5.0 statistical algorithm as implemented in Affymetrix Expression Console. Probe sets with >60% present calls in either of the two groups (OSE and CEPI) were selected for further analysis. After log2 transformation, the signal values of those probe sets were submitted to Statistical Analysis of Microarrays (SAM) for multiple testing correction where a 5.5% FDR was applied resulting in 7461 probe sets representing 5910 unique genes; annotations for probe sets were obtained from Affymetrix [34]. Genes were further filtered for absolute fold change > 1.5 between normal and cancer samples, resulting in 5144 differentially expressed (DE) genes. Further analysis of DE genes which were detected by exactly two probe sets revealed that ~2.5% of the probe sets reported contradictory results for a given gene; genes with contradictory probe sets were filtered out prior to network creation.

Correlation calculation

Pearson correlation between DE genes was calculated using the Mathematica (Wolfram Research 2010) correlation function. Since it is possible that the gene to gene correlation might be random, a significance test was conducted using expression signals from normal samples. For each Affymetrix probe set, expression signals from the 10 normal samples were shuffled among those 10 samples. That is, expression values were randomly reassigned to different samples, thus preserving the mean and standard deviation. By graphing the size of the largest connected subgraph for different values of r ranging from 0.70 to 0.99 [Supplemental Figure 1], it was found that networks of random signals can appear to be connected for values of $r < 0.85$. To minimize false positives, the absolute value of r was limited to values > 0.85. Baseline relationships between pairs of DE genes were established by correlating the expression of 12.8 million pairs of probes across 10 normal samples, and selecting pairs satisfying ($r > 0.85$; $df = 8$; $p < 0.0018$). This resulted in a total of 682391 highly correlated gene pairs. Random expression values can generate apparently highly correlated networks at $r < .85$ (Figure 4.1). Based on this analysis, r was set to .85 to minimize the possibility of unjustifiable correlations.

Gene enrichment

Lists of 350 inconsistent and 350 highly DE genes were subjected to genome ontology enrichment analysis by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang da, Sherman et al. 2009). The lists were analyzed using Functional Annotation Clustering. Biological process terms having $p < .05$ were considered significant.

Gene ontology selection

Cell cycle genes (n=577) (GO:0007049) and genes involved in the Regulation of Programmed Cell Death (n=892) (GO:0043067) were downloaded from QuickGO (Binns, Dimmer et al. 2009) after filtering for Taxon (human:9606), Evidence (Manual all), GO Identifier (GO: 0007049 or GO: 0043067 and their descendants), and Aspect (Biological Process). These genes formed the basis for Figures 1a and 1b, which were built using Cytoscape (Shannon, Markiel et al. 2003).

Graph construction and visualization

Cytoscape was used to visualize the correlation networks consisting of DE genes downloaded from QuickGO, by using a Prefuse Force Directed layout. The top 6 most highly correlated neighbors for each gene were connected to form a graph. Genes with only 1 neighbor are not shown.

CHAPTER 5

CONCLUSIONS

Throughout these studies, we were interested in investigating and explaining the possible causes of aberrant gene expression in ovarian cancer. The studies include analyses of interactions among genes, transcription factors and microRNAs, and explore how their dysregulation relates to cancer. Above all, our findings lead us to believe that it is difficult, if not impossible, to understand the breakdown of biological systems by simply measuring changes in expression of single genes or individual microRNAs. Instead, a deeper appreciation of the operation of a biological system as a network of interdependent elements requires some approximation, however crude, of the system itself, which can only be hinted at by lists of isolated, differentially expressed genes. And, while a detailed mechanistic model of even a single human cell is as yet intractable, some biologically meaningful approximation of that mechanism is needed to improve our current understanding. This dissertation basically addresses three questions: 1) to what extent do the predicted targets of miRNAs that differ by a single nucleotide in their seed region overlap; 2) what is the net effect on the expression of genes that are both predicted targets of upregulated microRNAs and negatively correlated targets of down-regulated transcriptional repressors; 3) what can we learn from **inconsistent** genes, that is, genes which change their relationship to other genes as the tissue they are expressed in becomes cancerous.

In the first study (Chapter 2), we showed how a single mutation in the 7-mer seed region of a microRNA can potentially affect the expression of hundreds of genes and therefore

influence a large number of biological processes. We found that while miRNAs with identical seeds share 88% of their mirSVR predicted targets, a single nt change reduced that overlap to about 18%. Subsequent mutations resulted in no additional loss of overlap. Surprisingly, even miRNAs whose seeds had no positions in common share about 16% of their targets. Due to the capacity of the 3'UTR of a targeted mRNA to accommodate a large number of miRNA binding sites, the mapping of seed sequences to functions that are mediated by their targeted genes, appears random. Random, in the sense that transcripts mediated by a miRNA with a novel seed are uninformed by a knowledge of transcripts mediated by other miRNAs with similar seeds. The high number of mutual targets among elements of individual miRNA families (miRNAs sharing identical seeds), accomplished both through seed equivalence and by having multiple binding sites on their targets, provides robustness (Siciliano, Garzilli et al. 2013) and process stability, but induces complexity (Lee, Risom et al. 2007; (Peter 2010). This may confound naive interpretations of miRNA transfection experiments; the high number of targets mediated by multiple miRNAs leads to complex interactions, including feedback loops (Lee, Risom et al. 2007) which are difficult to model or to tease apart experimentally (Tsang, Zhu et al. 2007). Thus, attribution of downstream mRNA expression variation, let alone specific biological functions, to a particular miRNA becomes computationally if not experimentally intractable. This underscores our claim that cancer must be approached as a system.

The idea that miRNA target determination is binding site dependent (rather than seed dependent) suggests a decoupling of miRNA seed structure from biological function; seeds are codes (Barbieri 2008). This observation questions assumptions of a causal link

between seed sequence and function, and instead proposes that seeds are biologically agnostic. This leads us to conclude that while miRNAs target mRNAs directly, they mediate biological functions indirectly. And while we do not deny that many miRNAs can be associated *a posteriori* with specific biological functions (Nam, Li et al. 2009; (Backes, Meese et al. 2010), we believe that that relationship is adventitious, and that the miRNAs only persist in regulating sets of mRNAs involved in evolutionarily beneficial biological functions.

In the second study (Chapter 3), we presented TOM, the transcriptional override model, in which we showed that even a simple feed-forward loop (FFL) motif can reveal unpredicted yet explanatory interactions among three genetic elements. Using TOM, we accounted for the indirect and unexpected upregulation of hundreds of predicted microRNA targets in ovarian cancer which might naively be anticipated to be down-regulated. We further showed that changes in expression among thousands of microRNA targets depend on a balance between the repressive effects of dozens of upregulated microRNAs, and the derepressive effects of over 100 down-regulated transcriptional repressors. The impact of these findings should not be underestimated; the possibility of microRNA mediated therapeutics will depend on an accurate assessment of both direct and indirect consequences of exogenous microRNA transfection.

The final study (Chapter 4) introduces the concept of **inconsistent** genes, which are derived from gene pairs which are highly correlated in normal tissue, but become dysregulated in diseased tissue. Surprisingly, most pairs of genes retain the same relationship in cancer as they did in normal samples; these **consistent** gene pairs

accounted for 94% of all significant relationships. And while **inconsistent** genes accounted for only about 6% of all correlated genes, we found many had previously been implicated in cancer and, even more interestingly, were involved in cell cycle and regulation of programmed cell death. We also found that there were fewer highly correlated genes pairs in cancer than in normal tissue, which we speculated might be accounted for by a loss of control in cancer, and/or the existence of multiple populations in the tumor. Further, we found that negatively correlated gene pairs in normal samples were more likely to become dysregulated (**inconsistent)** in cancer, and that those negatively correlated pairs were significantly more likely to be upregulated.

All three studies, while taking different approaches, emphasize the need to approach gene dysregulation in cancer from a systems perspective which requires the application of graph theoretical concepts and network related abstractions.

APPENDIX A

SUPPLEMENTARY DATA FOR CHAPTER 2

Table A.1: Fold-change for DE genes after MIR429 transfection. Sheet1: All 5229 differentially expressed genes between negative controls and transfected HEY cells. Sheet2: 874 predicted and down-regulated genes after transfection.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table A.2: Fold-change for DE genes after MIR200b transfection. Sheet1: All 6690 differentially expressed genes between negative controls and transfected HEY cells. Sheet2: 1006 predicted and down-regulated genes after transfection.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table A.3: Fold-change for DE genes after MIR141 transfection. Sheet1: All 7269 differentially expressed genes between negative controls and transfected HEY cells. Sheet2: 724 predicted and down-regulated genes after transfection.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table A.4: Fold-change for DE genes after MIR205 transfection. Sheet1: All 9596 differentially expressed genes between negative controls and transfected HEY cells. Sheet2: 593 predicted and down-regulated genes after transfection.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table A.5: Fold-change for DE genes after MIR419/M12 transfection. Sheet1: All 6225 differentially expressed genes between negative controls and transfected HEY cells. Sheet2: 135 predicted and down-regulated genes after transfection

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

APPENDIX B

SUPPLEMENTARY DATA FOR CHAPTER 3

Table B.1: Patient samples analyzed in this study. All tissues were collected according to previously published procedures (Bowen, Walker et al. 2009), following approved Institutional Review Board protocols from Northside Hospital (Atlanta) and Georgia Institute of Technology.

ID #	Age at time of surgery	Stage	Grade
183	66	III	2
336	63	Ic	3
369	52	IIIc	2
489	48	IV	3
528	66	IIIc	3
537	64	IIIa	2/3
542	61	IV	3
551	59	IIIc/IV	3
588	71	IIIc	2/3
606	54	IIIa	3
620	62	III/IV	3
434	41	N/A	N/A
437	54	N/A	N/A
440	50	N/A	N/A
452	51	N/A	N/A
470	44	N/A	N/A
475	63	N/A	N/A
482	44	N/A	N/A
567	77	N/A	N/A
665	84	N/A	N/A
541	41	N/A	N/A
552	41	N/A	N/A
563	66	N/A	N/A
783	52	N/A	N/A

Table B.2: Differentially expressed genes between OSE and EOC. Gene expression profiling identified 5910 significantly differentially expressed genes (mRNAs) between OSE and EOC. Of these, 2246 (38%) were significantly upregulated and 3664 (62%) downregulated in EOC.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table B.3: Down-regulated transcriptional repressors. Genes belonging to gene ontology categories GO:0045892, GO:0000122, GO:0010944, GO:0032088 and GO:0008156, relating to the negative-regulation-of-transcription, were downloaded from the European Bioinformatics Institute. Down-regulated repressor genes that were also predicted microRNA targets formed the 105 transcriptional repressor genes presented in this list.

Probeset_ID	Gene-Symbol	p-value	Fold-change
230141_at	ARID4A	0.00034	-4.29
209988_s_at	ASCL1	0.00612	-4.03
203232_s_at	ATXN1	0.00005	-2.45
204194_at	BACH1	0.00056	-3.65
236796_at	BACH2	0.00371	-3.07
202391_at	BASP1	0.03447	-2.44
201084_s_at	BCLAF1	0.00049	-3.24
205289_at	BMP2	0.00000	-8.73
207186_s_at	BPTF	0.00173	-6.64
224471_s_at	BTRC	0.00011	-2.84
212914_at	CBX7	0.00029	-2.58
235196_at	CDC73	0.00010	-3.72
209112_at	CDKN1B	0.00059	-4.40
204314_s_at	CREB1	0.02689	-2.09
209674_at	CRY1	0.00696	-1.77
201278_at	DAB2	0.00000	-22.47
200033_at	DDX5	0.00936	-3.69
204602_at	DKK1	0.00861	-2.08
204273_at	EDNRB	0.00336	-5.25
208669_s_at	EID1	0.00000	-13.74
212420_at	ELF1	0.03501	-2.80
203822_s_at	ELF2	0.00188	-3.77
225159_s_at	ELK4	0.00242	-2.10
225417_at	EPC1	0.00295	-1.74

Table B.3 (continued)

209455_at	FBXW11	0.00119	-2.17
202949_s_at	FHL2	0.00178	-2.83
226460_at	FNIP2	0.00038	-3.13
218031_s_at	FOXN3	0.00383	-2.28
224891_at	FOXO3	0.00195	-2.23
235444_at	FOXP1	0.00000	-5.64
210002_at	GATA6	0.00000	-56.60
225884_s_at	GZF1	0.00700	-2.74
228813_at	HDAC4	0.00098	-2.51
226648_at	HIF1AN	0.00146	-1.96
219269_at	HMBOX1	0.01083	-2.39
210338_s_at	HSPA8	0.00023	-5.40
201565_s_at	ID2	0.01252	-2.02
209292_at	ID4	0.00001	-16.97
206332_s_at	IFI16	0.00000	-7.22
225798_at	JAZF1	0.00009	-2.70
218486_at	KLF11	0.00043	-2.37
221841_s_at	KLF4	0.00021	-5.16
222561_at	LANCL2	0.00009	-2.13
209348_s_at	MAF	0.00000	-10.56
236814_at	MDM4	0.01674	-2.34
212535_at	MEF2A	0.00002	-4.99
209200_at	MEF2C	0.00000	-5.66
212251_at	MTDH	0.00219	-4.70
219363_s_at	MTERFD1	0.00816	-2.04
212993_at	NACC2	0.00055	-4.15
200854_at	NCOR1	0.00233	-1.77
213012_at	NEDD4	0.00008	-1.73
213032_at	NFIB	0.00147	-5.00
203574_at	NFIL3	0.00582	-2.27
209239_at	NFKB1	0.00116	-1.66
223439_at	NKAP	0.01086	-1.99
209706_at	NKX3-1	0.00000	-20.57
206645_s_at	NR0B1	0.00000	-6.22
212594_at	PDCD4	0.00000	-4.50
210296_s_at	PEX2	0.00184	-6.97
228469_at	PPID	0.00001	-2.87
235764_at	PRDM5	0.00000	-7.93
223275_at	PRMT6	0.00698	-2.27
219485_s_at	PSMD10	0.02809	-3.19
207785_s_at	RBPJ	0.00007	-3.50

Table B.3 (continued)

204633_s_at	RPS6KA5	0.00353	-1.98
222540_s_at	RSF1	0.06553	-2.57
203408_s_at	SATB1	0.00003	-6.71
40189_at	SET	0.00288	-3.52
223122_s_at	SFRP2	0.00001	-18.40
218878_s_at	SIRT1	0.00108	-2.71
203076_s_at	SMAD2	0.00224	-2.20
202527_s_at	SMAD4	0.00503	-3.04
206542_s_at	SMARCA2	0.00005	-8.79
211988_at	SMARCE1	0.00006	-3.90
213139_at	SNAI2	0.00010	-4.91
202864_s_at	SP100	0.00135	-4.37
232529_at	SP3	0.00021	-5.91
201996_s_at	SPEN	0.01748	-3.31
201023_at	TAF7	0.00019	-2.57
226037_s_at	TAF9B	0.00000	-6.58
235890_at	TBL1XR1	0.00204	-2.42
225544_at	TBX3	0.00178	-3.21
227705_at	TCEAL7	0.00000	-19.44
204931_at	TCF21	0.00043	-3.93
201730_s_at	TPR	0.00958	-4.40
223393_s_at	TSHZ3	0.00000	-5.95
204771_s_at	TTF1	0.01342	-1.90
208760_at	UBE2I	0.00002	-5.03
220746_s_at	UIMC1	0.00029	-2.84
223118_s_at	USP47	0.00015	-4.99
213425_at	WNT5A	0.00000	-7.21
206067_s_at	WT1	0.00146	-3.43
208643_s_at	XRCC5	0.00082	-2.37
224718_at	YY1	0.00076	-2.75
212764_at	ZEB1	0.00019	-4.74
228333_at	ZEB2	0.00020	-2.90
219778_at	ZFPM2	0.00000	-33.39
223213_s_at	ZHX1	0.00380	-2.86
226015_at	ZNF12	0.00036	-4.27
228545_at	ZNF148	0.00001	-6.64
203247_s_at	ZNF24	0.00002	-4.38
218401_s_at	ZNF281	0.00033	-3.48
225539_at	ZNF295	0.00051	-4.22
219266_at	ZNF350	0.00134	-2.19

Table B.4: Upregulated mutual targets of upregulated miRNAs and down-regulated transcriptional repressors. List of 843 TRANSFAC identified targets of 10 repressor genes (Table 3.3) that are also targets of one or more of the 31 miRNAs upregulated in EOC.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

APPENDIX C

SUPPLEMENTARY DATA FOR CHAPTER 4

Table C.1: Top 5070 differentially expressed genes. This table shows the 5070 significantly differentially and highly correlated genes, with the fold change (FC) used in these analyses.

Please see table on our website:
http://www.mcdonaldlab.biology.gatech.edu/dissertations/hill.htm

Table C.2: The top 350 genes involved in the greatest number of inconsistent gene-pair interactions. This table shows the 350 genes involved in the greatest number of inconsistent pairings, along with the number of inconsistent interactions. These genes were submitted to DAVID for enrichment.

Gene symbol	Inconsistent interactions	Gene symbol	Inconsistent interactions	Gene symbol	Inconsistent interactions
VRK1	650	XPR1	645	COL1A2	612
HELLS	576	TET1	574	CSTF3	558
SVIL	528	PTPRM	506	RMI1	495
LRIG1	476	SLCO3A1	464	GPSM2	463
SLC35E3	447	CHD7	438	HMGB3	430
COL5A2	419	EZR	398	FAM72A	396
FAM171A1	395	APOLD1	388	GYG1	367
C14orf143	365	GUCY1B3	352	PPIC	346
TFDP1	323	FGFR1OP	322	COL6A3	321
FCHSD2	310	ANP32E	308	TPM1	307
ZWILCH	294	BEND3	274	ASPM	272
HNRNPA1	271	SLC4A7	268	MAD2L1	263
FNDC3B	257	RPP40	257	PCDHB2	245
CTHRC1	238	DYRK2	236	PHF16	235
BTN3A2	232	EIF2C2	231	NETO2	231
NPAS2	230	THBS2	228	ATAD2	225
SASS6	225	SPIN4	225	LARP4	222
CCNA2	221	JPH1	221	SIX4	219

Table C.2 (continued)

GAS5	212	TMCC3	211	CCNB1	210
ENAH	210	DNA2	209	FANCD2	209
SCARNA15	208	SPON1	208	TMEM48	208
TMSB15A	205	MRPL3	199	GALNT2	194
SELENBP1	194	WWOX	189	NACA	188
SMG7	188	CHEK1	187	TMEM189	182
CENPK	178	STMN1	177	HERC6	174
MYO10	171	RPIA	171	WHSC1	171
RAD52	169	ZNF738	169	TARBP1	165
RBPMS	162	COL15A1	161	KIAA1524	161
FAM119A	159	TBCCD1	159	SRPK1	155
SMARCC1	154	ADH1B	153	CD59	151
TRIB2	148	JAG1	145	RNF144B	143
RUNX1	143	TIA1	142	PCDH7	140
PXDN	139	BCL2L11	136	PTPRG	136
CKS1B	132	MEIS1	132	ECT2	130
PFDN4	129	SLC25A33	129	EFEMP1	128
HTATIP2	128	FLNC	125	RAB4A	125
C21orf63	124	GEN1	121	GCLC	120
KIAA0101	118	SORBS1	117	TOPBP1	116
CDC16	115	ASAP2	113	HIST1H2AG	112
ESYT2	111	DLG2	110	XRCC6	109
NEAT1	105	BICD1	104	FNIP2	104
BARD1	103	CENPF	102	ZNF439	102
MTHFD2	100	RAP1GAP2	100	IGDCC4	99
CASP2	98	EPSTI1	98	KIAA0907	98
MNS1	96	EPB41L5	95	ADCY7	94
TMEM178	94	SOAT1	93	GPRC5B	92
ZDHHC9	91	GINS1	89	CDH6	88
RAD51AP1	88	C5orf34	87	FAM108C1	87
TSGA14	87	C15orf27	86	GNB5	85
CSNK2A1	84	PTPRO	84	PAIP1	83
CTPS2	82	PRDXDD1P	82	ASAH1	81
PNO1	81	PRKCI	81	WDHD1	81
HOXA3	80	B4GALT5	79	CDK1	79
SMC2	77	COL4A1	76	SMAD1	76

Table C.2 (continued)

PELI1	75	PLK4	75	PUS7	75
CHSY1	74	TTF2	73	ZNF585A	73
TMEM37	72	TTC30A	72	DACH1	71
ITGB8	71	HMMR	70	KRT19	70
TLR1	70	CD93	69	SETD4	69
TAGAP	69	VANGL1	69	ANKRD12	68
GATA5	68	PIAS3	68	SMC4	68
CFC1	67	CHD2	67	EZH2	67
LUZP1	67	PHLDA1	67	RNF44	67
CHCHD7	66	LTBP1	66	TMCO6	66
C11orf9	65	CASP1	65	N4BP2L2	65
SAMD9L	65	TIGD1	65	TMEM154	65
TST	65	UNG	65	GNAS	63
LRRK2	63	MAML2	63	R3HDM1	63
SLC16A1	63	HABP4	62	SMCHD1	62
ZWINT	62	C13orf34	61	C9orf72	61
DLGAP5	61	IFIT2	61	POMGNT1	61
QKI	61	R3HCC1	61	TIMELESS	61
TLR2	61	ASAP1	60	ELF1	60
HEY2	60	MASTL	60	MIR21	60
SLC7A11	60	CACNB3	59	COPZ1	59
OSBPL3	59	SCD	59	TRIP13	59
ZNF682	59	CHMP2B	58	LRRN4	58
SNHG12	57	FEN1	56	PHF17	56
PRC1	56	CD274	55	EIF3M	55
MAP1LC3B	55	SEMA6A	55	BNIP3L	54
CD53	54	KAZ	54	MDC1	54
PLEKHH1	54	XK	53	CBX3	52
FCGR3B	52	WTAP	52	AURKA	51
CD83	51	CDV3	51	FGFR1OP2	51
HCP5	51	KIF11	51	RRAGC	51
SEC14L1	51	STK17A	51	BBX	50
CEP55	50	NFIA	50	RILPL2	50
CASP4	49	CREBL2	49	FAM46A	49
MMP28	49	MYEF2	49	NEBL	49
PEA15	49	PRKCB	49	TANK	49

Table C.2 (continued)

COMMD5	48	FOXP1	48	MCM2	48
NEK6	48	SORL1	48	CNOT2	47
FNDC1	47	HERC5	47	PTX3	47
RHOQ	47	METT10D	46	PLLP	46
SPAG9	46	USP3	46	VCAN	46
ZNF292	46	DENND1B	45	EIF5	45
GPR4	45	LITAF	45	PDE4B	45
PLCG2	45	RAB34	45	SP3	45
ST3GAL6	45	ZFP82	45	BECN1	44
CAP1	44	EIF5B	44	ERBB2IP	44
ESRP1	44	VEZF1	44	ARL6IP5	43
GPC2	43	LAP3	43	MPZL2	43
NF1	43	NFIB	43	SLC15A4	43
SP100	43	TICAM2	43	TNFAIP3	43
TPP1	43	BRD7	42	COL4A3BP	42
DSCC1	42	IGF2BP2	42	JMJD1C	42
LCORL	42	LHFPL2	42	MIR17HG	42
NPL	42	SFMBT1	42	SKAP2	42
SLC35A2	42	THBS1	42	CD47	41
CLCN3	41	EIF5A	41	HMGA2	41
LARP4B	41	MELK	41	NFIL3	41
NPEPPS	41	NPY	41	PDLIM5	41
RNF13	41	RORC	41	SFTPD	41
SLC1A1	41	SMAD2	41	TMED5	41
AKR7A2	40	C22orf32	40	CAMK2D	40
DDX3X	40	IQGAP1	40	KLHL9	40
PELO	40	PION	40		

Table C.3: The top 350 most significantly differentially expressed genes between control and cancer samples. This table shows the 350 most differentially expressed genes along with their fold change (FC). These genes were submitted to DAVID for enrichment.

Gene Symbol	Fold change	Gene Symbol	Fold change	Gene Symbol	Fold change
ITLN1	-9.29	ALDH1A2	-8.49	LHX2	-8.017
SFRP1	-7.76	BNC1	-7.45	CHRDL1	-7.317
LHX9	-7.25	BCHE	-7.13	PDGFD	-7.062
ABCA8	-7.01	NELL2	-6.99	SOX17	6.922
REEP1	-6.89	ADH1B	-6.73	EFEMP1	-6.725
AOX1	-6.68	C13orf36	-6.55	OGN	-6.544
C1orf168	-6.48	ANXA8	-6.45	SLC4A4	-6.391
RARRES1	-6.32	AQP9	-6.31	TCEAL2	-6.267
ARHGAP18	-6.24	SPOCK1	-6.23	CD24	6.169
RAB27B	-6.05	C4orf49	-5.94	DSC3	-5.902
TFPI2	-5.89	TXNIP	-5.86	DPYD	-5.844
GATA6	-5.82	FAM153A	-5.78	DOCK11	-5.764
PDE7B	-5.73	RGS4	-5.72	WNT2B	-5.694
PRG4	-5.61	MUM1L1	-5.59	KLHL14	5.433
C8orf84	-5.43	HOXD8	-5.42	FAM70A	-5.417
C21orf62	-5.40	EZR	-5.39	SEMA3C	-5.333
DFNA5	-5.33	SLC4A11	5.27	CAV2	-5.272
ARX	-5.27	PSAT1	5.23	SGCG	-5.225
CRNDE	-5.21	PCDH17	-5.18	EHF	5.176
PRKAR2B	-5.16	PDE8B	-5.14	DCN	-5.135
GAS1	-5.09	LXN	-5.08	CYBRD1	-5.074
ZFPM2	-5.06	FGF13	-5.04	PGR	-5.018
STK31	-5.01	TNFSF13B	-4.99	MFSD4	-4.948
C10orf58	-4.93	MAL	4.93	SLC39A8	-4.913
GADL1	-4.90	TSPAN5	-4.89	MTUS1	-4.887
MUC1	4.88	RERG	-4.87	CDH6	4.857
FLRT2	-4.85	MNDA	-4.81	ALDH1A3	-4.808
EMB	-4.76	PTGER4	-4.74	SLC2A1	4.742
KITLG	-4.74	S100A2	4.72	CSGALNACT1	-4.672
ABHD12B	-4.61	LGALS2	-4.61	FLRT3	-4.577
ZIC1	4.56	MMP28	-4.56	SPON1	4.556
PLSCR4	-4.52	MEIS2	-4.51	NEFH	-4.506
BNC2	-4.50	C2orf40	-4.50	DAB2	-4.490

Table C.3 (continued)

OMD	-4.48	GHR	-4.47	CRIM1	-4.465
PROCR	-4.45	LGALS8	-4.45	NAP1L3	-4.437
PAPSS2	-4.43	SYT1	-4.42	CELF2	-4.418
GPM6A	-4.42	CYP3A5	-4.39	SYT4	-4.375
NKX3-1	-4.36	GABRG1	-4.36	C4orf31	-4.350
RYR2	-4.33	PRKAR1A	-4.32	PTGIS	-4.313
DSE	-4.31	PTGER3	-4.29	NUSAP1	4.288
SNCA	-4.28	TCEAL7	-4.28	ECM2	-4.279
ZWINT	4.27	AIF1L	4.27	PCDH9	-4.252
PITPNC1	-4.25	PPAP2A	-4.24	CENPF	4.241
STON2	4.23	SCG5	-4.22	KCNJ8	-4.216
MECOM	4.21	FRY	-4.21	KCNT2	-4.202
SFRP2	-4.20	MPZL2	4.20	USP53	-4.183
DAPK1	-4.17	CDCA3	4.16	ST7L	-4.158
ANKRD29	-4.16	LIFR	-4.15	MST4	-4.142
EEF1A1	-4.14	IRX3	-4.13	CALB2	-4.116
CDKN2A	4.11	LRRN4	-4.11	SEMA5A	-4.109
SCD5	-4.10	CTHRC1	4.09	ID4	-4.085
VIM	-4.08	AURKA	4.08	GOLM1	-4.079
CNTN4	-4.07	KAL1	-4.06	COL3A1	-4.054
LIPA	-4.05	CLIP4	-4.05	MBNL3	-4.041
LYPLAL1	-4.04	S100A8	-4.03	GPNMB	-4.031
EFNB3	-4.03	OLFML1	-4.03	TIAM1	-4.026
NBLA00301	-4.02	DTNA	-4.02	RTN1	-4.013
HSD17B2	-4.00	CFL2	-3.99	SLC16A1	-3.985
SCGB2A1	3.98	CAV1	-3.98	AG2	3.972
ABI3BP	-3.94	CFI	-3.94	MCOLN3	-3.937
PAX8	3.91	DDR2	-3.91	MET	-3.906
NPY1R	-3.90	HPGD	-3.89	RNF128	-3.884
NAV3	-3.87	METTL7A	-3.87	THBD	-3.868
WFDC2	3.86	CCDC80	-3.86	SGCE	-3.849
GFPT2	-3.85	CNTNAP4	-3.85	LAMA4	-3.833
CXCR4	3.82	TRIM4	-3.81	SFN	3.796
ARRDC4	-3.79	PRSS35	-3.79	CLDN1	-3.790
SNX13	-3.79	RABEP1	-3.79	GSTM3	-3.784
EID1	-3.78	TPD52L1	-3.78	PPM1K	-3.775
PHACTR2	-3.76	PLCL2	-3.76	SULT1C2	3.746
RBMS3	-3.74	TFPI	-3.74	DPP10	-3.733
PROS1	-3.72	SLC41A2	-3.71	AKT3	-3.705

Table C.3 (continued)

LIMA1	-3.70	SYBU	-3.69	ADH5	-3.692
HMMR	3.68	CHGB	-3.68	SORBS2	-3.678
NDN	-3.67	NBEA	-3.67	BAMBI	-3.671
KIAA0101	3.67	PROM2	3.67	RRM2	3.666
ANXA1	-3.66	VLDLR	-3.66	TMTC1	3.659
UGP2	-3.66	COBLL1	-3.66	SDC2	-3.655
SNX7	-3.65	C5orf41	-3.65	PTPLA	-3.645
PPM1E	-3.64	C13orf33	-3.64	CFH	-3.621
H3F3B	-3.62	PMP22	-3.62	GULP1	-3.617
LSAMP	-3.61	SERINC1	-3.60	MSRB3	-3.592
CP	3.59	C3	-3.59	CAST	-3.590
NAP1L2	-3.59	B2M	-3.59	IL18	-3.566
MTHFD2	3.57	CLK1	-3.56	GALC	-3.561
CARD16	-3.56	AKAP11	-3.55	KIAA1244	-3.552
PLEKHH2	-3.55	FBXO9	-3.54	HPSE	-3.538
FGF18	3.54	SPOPL	-3.53	RAB38	-3.531
CTNNAL1	-3.52	EPCAM	3.52	GIPC2	-3.502
UBL3	-3.50	SCP2	-3.48	NR3C1	-3.479
PHLPP2	-3.48	NR3C2	-3.47	NXPH2	-3.464
FHL1	-3.46	GNB4	-3.46	FAM89A	-3.455
SQRDL	-3.45	RUFY3	-3.45	HMGA1	3.444
RSPO1	-3.43	TFAP2A	3.43	DMD	-3.428
GRIN2A	-3.42	PPP1R3B	-3.42	EIF3M	-3.417
ANXA4	-3.42	VCAN	3.41	MGC24103	-3.411
NTN4	-3.41	CARD16	-3.41	CCNE1	3.408
TMOD1	-3.41	CLDN15	-3.40	ARHGAP28	-3.403
MAF	-3.40	C1orf186	3.39	PKHD1L1	-3.390
CFH	-3.39	CNRIP1	-3.38	MCTP2	-3.380
SDPR	-3.38	OSTM1	-3.37	FAM13C	-3.368
GCOM1	-3.37	ANXA3	-3.37	PEX5L	-3.365
C8orf4	3.36	C9orf5	-3.36	PCDH7	3.348
PLCE1	-3.35	RAB31	-3.34	CEP55	3.334
FBXO3	-3.33	NRXN3	-3.33	ITPR2	-3.329
NLGN4X	-3.32	ESD	-3.32	SMAD9	-3.310
LEPROTL1	-3.31	GPR137B	-3.30	CKS2	3.294
CXXC5	3.29	ME1	-3.29	PKD2	-3.284
CPE	-3.28	MELK	3.28	TGFB2	-3.281
TRPC1	-3.28	DDX17	-3.28	CMAH	-3.276
SEL1L2	-3.28	CDK1	3.27	LIMS3	3.260

Table C.3 (continued)

PHLDB2	-3.25	KLHDC8A	-3.25	CYP39A1	-3.248
FNDC3A	-3.25	RCHY1	-3.25	NEK2	3.240
ELF3	3.24	FGL2	-3.23	CLEC4M	-3.233
MAOA	-3.23	TRPA1	-3.23	NT5DC1	-3.226
FRAS1	-3.22	CGNL1	-3.21	SIX1	3.213
HNRPDL	-3.21	SLC44A1	-3.20		

REFERENCES

Abdel-Fatah, T. M., R. Russell, et al. (2014). "Genomic and protein expression analysis reveals flap endonuclease 1 (FEN1) as a key biomarker in breast and ovarian cancer." Mol Oncol **8**(7): 1326-1338.

Affymetrix "Affymetrix® Expression Console™ Software."

Ahmed, J., T. Meinel, et al. (2011). "CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge." Nucleic Acids Res **39**(Database issue): D960-967.

Ahn, A. C., M. Tewari, et al. (2006). "The limits of reductionism in medicine: could systems biology offer an alternative?" PLoS Med **3**(6): e208.

Allocco, D. J., I. S. Kohane, et al. (2004). "Quantifying the relationship between co-expression, co-regulation and gene function." BMC Bioinformatics **5**: 18.

Alvarez-Garcia, I. and E. A. Miska (2005). "MicroRNA functions in animal development and human disease." Development **132**(21): 4653-4662.

Amar, D., H. Safer, et al. (2013). "Dissection of regulatory networks that are altered in disease via differential co-expression." PLoS Comput Biol **9**(3): e1002955.

Ariyoshi, M. and J. W. Schwabe (2003). "A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling." Genes Dev **17**(15): 1909-1920.

Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Auranen, A., H. Song, et al. (2005). "Polymorphisms in DNA repair genes and epithelial ovarian cancer risk." International Journal of Cancer **117**(4): 611-618.

Backes, C., E. Meese, et al. (2010). "A dictionary on microRNAs and their putative target pathways." Nucleic Acids Res **38**(13): 4476-4486.

Barabasi, A. L. (2007). "Network medicine--from obesity to the "diseasome"." New England Journal of Medicine **357**(4): 404-407.

Barabasi, A. L., N. Gulbahce, et al. (2011). "Network medicine: a network-based approach to human disease." Nat Rev Genet **12**(1): 56-68.

Barabasi, A. L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." Nat Rev Genet **5**(2): 101-113.

Barbieri, M. (2008). "Biosemiotics: a new understanding of life." Naturwissenschaften **95**(7): 577-599.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-297.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-233.

Bast, R. C., Jr. (2011). "Molecular approaches to personalizing management of ovarian cancer." Ann Oncol **22 Suppl 8**: viii5-viii15.

Bertoni, F., A. M. Codegoni, et al. (1999). "CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers." Genes, Chromosomes and Cancer **26**(2): 176-180.

Betel, D., A. Koppal, et al. (2010). "Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites." Genome Biol **11**(8): R90.

Binns, D., E. Dimmer, et al. (2009). "QuickGO: a web-based tool for Gene Ontology searching." Bioinformatics **25**(22): 3045-3046.

Bonavida, B. and S. Baritaki (2011). "The novel role of Yin Yang 1 in the regulation of epithelial to mesenchymal transition in cancer via the dysregulated NF-kappaB/Snail/YY1/RKIP/PTEN Circuitry." Crit Rev Oncog **16**(3-4): 211-226.

Bowen, N. J., L. D. Walker, et al. (2009). "Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells." Bmc Medical Genomics **2**.

Brendle, A., A. Brandt, et al. (2009). "Single nucleotide polymorphisms in chromosomal instability genes and risk and clinical outcome of breast cancer: a Swedish prospective case-control study." Eur J Cancer **45**(3): 435-442.

Buick, R. N., R. Pullano, et al. (1985). "Comparative properties of five human ovarian adenocarcinoma cell lines." Cancer Res **45**(8): 3668-3676.

Cancer Genome Atlas Research, N. (2011). "Integrated genomic analyses of ovarian carcinoma." Nature **474**(7353): 609-615.

Cordero, D., X. Sole, et al. (2014). "Large differences in global transcriptional regulatory programs of normal and tumor colon cells." BMC Cancer **14**: 708.

Dachs, G. U., G. J. Dougherty, et al. (1997). "Targeting gene therapy to cancer: a review." Oncology Research **9**(6-7): 313-325.

Deng, C. X. (2006). "BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution." Nucleic Acids Res **34**(5): 1416-1426.

Di Leva, G. and C. M. Croce (2013). "miRNA profiling of cancer." Curr Opin Genet Dev **23**(1): 3-11.

Djebbari, A. and J. Quackenbush (2008). "Seeded Bayesian Networks: constructing genetic networks from microarray data." BMC Syst Biol **2**: 57.

Ehrlichova, M., B. Mohelnikova-Duchonova, et al. (2013). "The association of taxane resistance genes with the clinical course of ovarian carcinoma." Genomics **102**(2): 96-101.

Eijssen, L. M., M. Jaillard, et al. (2013). "User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org." Nucleic Acids Res **41**(Web Server issue): W71-76.

Emmanuel, C., N. Gava, et al. (2011). "Comparison of expression profiles in ovarian epithelium in vivo and ovarian cancer identifies novel candidate genes involved in disease pathogenesis." PLoS ONE **6**(3): e17617.

Forbes, S. A., N. Bindal, et al. (2011). "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." Nucleic Acids Res **39**(Database issue): D945-950.

Ford, H. L. and A. B. Pardee (1999). "Cancer and the cell cycle." Journal of Cellular Biochemistry **Suppl 32-33**: 166-172.

Friedman, R. C., K. K. H. Farh, et al. (2009). "Most mammalian mRNAs are conserved targets of microRNAs." Genome Research **19**(1): 92-105.

Garofalo, M. and C. M. Croce (2011). "microRNAs: Master regulators as potential therapeutics in cancer." Annu Rev Pharmacol Toxicol **51**: 25-43.

Giroux, V., J. Iovanna, et al. (2006). "Probing the human kinome for kinases involved in pancreatic cancer cell survival and gemcitabine resistance." FASEB Journal **20**(12): 1982-1991.

Gordon, S., G. Akopyan, et al. (2006). "Transcription factor YY1: structure, function, and therapeutic implications in cancer biology." Oncogene **25**(8): 1125-1142.

Gourley, C., A. J. Paige, et al. (2005). "WWOX mRNA expression profile in epithelial ovarian cancer supports the role of WWOX variant 1 as a tumour suppressor, although the role of variant 4 remains unclear." International Journal of Oncology **26**(6): 1681-1689.

Gurtan, A. M. and P. A. Sharp (2013). "The Role of miRNAs in Regulating Gene Expression Networks." J Mol Biol.

Gyorffy, B., M. Dietel, et al. (2008). "A snapshot of microarray-generated gene expression signatures associated with ovarian carcinoma." Int J Gynecol Cancer **18**(6): 1215-1233.

Ha, T. Y. (2011). "MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease." Immune Netw **11**(3): 135-154.

Haber, D. A. and J. Settleman (2007). "Cancer: drivers and passengers." Nature **446**(7132): 145-146.

Hamming, R. W. (1950). "Error Detecting and Error Correcting Codes." Bell System Technical Journal **29**(2): 147-160.

Harris, S. L. and A. J. Levine (2005). "The p53 pathway: positive and negative feedback loops." Oncogene **24**(17): 2899-2908.

Helleday, T., E. Petermann, et al. (2008). "DNA repair pathways as targets for cancer therapy." Nat Rev Cancer **8**(3): 193-204.

Helleman, J., M. P. Jansen, et al. (2006). "Molecular profiling of platinum resistant ovarian cancer." International Journal of Cancer **118**(8): 1963-1971.

Hill, C. G., L. V. Matyunina, et al. (2014). "Transcriptional override: a regulatory network model of indirect responses to modulations in microRNA expression." BMC Syst Biol **8**(36).

Hornberg, J. J., F. J. Bruggeman, et al. (2006). "Cancer: a Systems Biology disease." Biosystems **83**(2-3): 81-90.

Hsu, C. H., K. L. Peng, et al. (2012). "TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases." Cell Rep **2**(3): 568-579.

Huang da, W., B. T. Sherman, et al. (2009). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res **37**(1): 1-13.

Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Huang, R. Y., G. B. Chen, et al. (2012). "Histotype-specific copy-number alterations in ovarian cancer." Bmc Medical Genomics **5**.

Huntzinger, E. and E. Izaurralde (2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay." Nat Rev Genet **12**(2): 99-110.

Ideker, T., T. Galitski, et al. (2001). "A new approach to decoding life: systems biology." Annu Rev Genomics Hum Genet **2**: 343-372.

Jabbari, N., A. N. Reavis, et al. (2014). "Sequence variation among members of the miR-200 microRNA family is correlated with variation in the ability to induce hallmarks of mesenchymal-epithelial transition in ovarian cancer cells." J Ovarian Res **7**(1): 12.

Jia, D., S. M. Hasso, et al. (2013). "Transcriptional repression of VEGF by ZNF24: mechanistic studies and vascular consequences in vivo." Blood **121**(4): 707-715.

John, B., A. J. Enright, et al. (2004). "Human MicroRNA targets." PLoS Biol **2**(11): e363.

Kang, J., W. Zheng, et al. (2011). "Use of Bayesian networks to dissect the complexity of genetic disease: application to the Genetic Analysis Workshop 17 simulated data." BMC Proc **5 Suppl 9**: S37.

Kar, G., A. Gursoy, et al. (2009). "Human cancer protein-protein interaction network: a structural perspective." PLoS Comput Biol **5**(12): e1000601.

Karlan, B. Y., J. Dering, et al. (2014). "POSTN/TGFBI-associated stromal signature predicts poor prognosis in serous epithelial ovarian cancer." Gynecologic Oncology **132**(2): 334-342.

Karppinen, S. M., H. Erkko, et al. (2006). "Identification of a common polymorphism in the TopBP1 gene associated with hereditary susceptibility to breast and ovarian cancer." Eur J Cancer **42**(15): 2647-2652.

Kasprzyk, A. (2011). "BioMart: driving a paradigm change in biological data management." Database (Oxford) **2011**: bar049.

Kerr, J. F., A. H. Wyllie, et al. (1972). "Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics." Br J Cancer **26**(4): 239-257.

Khanna, A., O. Kauko, et al. (2013). "Chk1 targeting reactivates PP2A tumor suppressor activity in cancer cells." Cancer Research **73**(22): 6757-6769.

Kim, D., R. Li, et al. (2014). "Knowledge-driven genomic interactions: an application in ovarian cancer." BioData Min **7**: 20.

Kitamuro, T., K. Takahashi, et al. (2003). "Bach1 functions as a hypoxia-inducible repressor for the heme oxygenase-1 gene in human cells." J Biol Chem **278**(11): 9125-9133.

Kozomara, A. and S. Griffiths-Jones (2014). "miRBase: annotating high confidence microRNAs using deep sequencing data." Nucleic Acids Res **42**(Database issue): D68-73.

Krek, A., D. Grun, et al. (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.

Kumtepe, Y., Z. Halici, et al. (2013). "High serum HTATIP2/TIP30 level in serous ovarian cancer as prognostic or diagnostic marker." European Journal of Medical Research **18**.

Laubenbacher, R. and A. S. Jarrah (2009). "Algebraic models of biochemical networks." Methods Enzymol **467**: 163-196.

Laubenbacher, R. and B. Stigler (2004). "A computational algebra approach to the reverse engineering of gene regulatory networks." J Theor Biol **229**(4): 523-537.

Laufer, C., B. Fischer, et al. (2013). "Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping." Nat Methods **10**(5): 427-431.

Lee, C. T., T. Risom, et al. (2007). "Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny." DNA Cell Biol **26**(4): 209-218.

Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-854.

Lewis, A. G., J. Flanagan, et al. (2005). "Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer." Breast Cancer Res **7**(6): R1005-1016.

Lewis, B. P., I. H. Shih, et al. (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-798.

Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-773.

Liu, H., D. Yue, et al. (2010). "Improving performance of mammalian microRNA target prediction." BMC Bioinformatics **11**: 476.

Madhamshettiwar, P. B., S. R. Maetschke, et al. (2012). "Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets." Genome Med **4**(5): 41.

Mansson, R., P. Tsapogas, et al. (2004). "Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor." J Biol Chem **279**(17): 17905-17913.

Matkovich, S. J., Y. Hu, et al. (2013). "Regulation of Cardiac MicroRNAs by Cardiac MicroRNAs." Circ Res **113**(1): 62-71.

Matys, V., O. V. Kel-Margoulis, et al. (2006). "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes." Nucleic Acids Res **34**(Database issue): D108-110.

Miles, G. D., M. Seiler, et al. (2012). "Identifying microRNA/mRNA dysregulations in ovarian cancer." BMC Res Notes **5**: 164.

Molitor, T. P. and P. Traktman (2013). "Molecular genetic analysis of VRK1 in mammary epithelial cells: depletion slows proliferation in vitro and tumor growth and metastasis in vivo." Oncogenesis **2**: e48.

Moolthiya, W. and P. Yuenyao (2009). "The risk of malignancy index (RMI) in diagnosis of ovarian malignancy." Asian Pac J Cancer Prev **10**(5): 865-868.

Moreau, Y. and L. C. Tranchevent (2012). "Computational tools for prioritizing candidate genes: boosting disease gene discovery." Nat Rev Genet **13**(8): 523-536.

Nagaraj, S. and A. Reverter (2011). "A Boolean-based systems biology approach to predict novel genes associated with cancer: Application to colorectal cancer." BMC Syst Biol **5**(1): 35.

Nam, S., M. Li, et al. (2009). "MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression." Nucleic acids research **37**(suppl 2): W356-W362.

Nana-Sinkam, S. P. and C. M. Croce (2010). "MicroRNA dysregulation in cancer: opportunities for the development of microRNA-based drugs." IDrugs **13**(12): 843-846.

Napoli, C., C. Lemieux, et al. (1990). "Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans." Plant Cell **2**(4): 279-289.

Nazarov, P. V., S. E. Reinsbach, et al. (2013). "Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function." Nucleic Acids Res.

Nikolova, D. N., N. Doganov, et al. (2009). "Genome-wide gene expression profiles of ovarian carcinoma: Identification of molecular targets for the treatment of ovarian carcinoma." Molecular Medicine Reports **2**(3): 365-384.

Ouyang, G., L. Yao, et al. (2009). "Genistein induces G2/M cell cycle arrest and apoptosis of human ovarian cancer cells via activation of DNA damage checkpoint pathways." Cell Biol Int **33**(12): 1237-1244.

Oyake, T., K. Itoh, et al. (1996). "Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site." Molecular and Cellular Biology **16**(11): 6083-6095.

Park, P. E., J. Y. Jeong, et al. (2013). "MAD2 Expression in Ovarian Carcinoma: Different Expression Patterns and Levels among Various Types of Ovarian Carcinoma and Its Prognostic Significance in High-Grade Serous Carcinoma." Korean J Pathol **47**(5): 418-425.

Pe'er, D. and N. Hacohen (2011). "Principles and strategies for developing network models in cancer." Cell **144**(6): 864-873.

Peter, M. E. (2010). "Targeting of mRNAs by multiple miRNAs: the next step." Oncogene **29**(15): 2161-2164.

Pomerening, J. R. (2009). "Positive-feedback loops in cell cycle progression." FEBS Letters **583**(21): 3388-3396.

Ramakrishna, M., L. H. Williams, et al. (2010). "Identification of candidate growth promoting genes in ovarian cancer through integrated copy number and expression analysis." PLoS One **5**(4): e9983.

Remenyi, A., H. R. Scholer, et al. (2004). "Combinatorial control of gene expression." Nat Struct Mol Biol **11**(9): 812-815.

Ritchie, W., J. E. Rasko, et al. (2013). "MicroRNA target prediction and validation." Adv Exp Med Biol **774**: 39-53.

Roth, R. B., P. Hevezi, et al. (2006). "Gene expression analyses reveal molecular relationships among 20 regions of the human CNS." Neurogenetics **7**(2): 67-80.

Schafer, J. and K. Strimmer (2005). "An empirical Bayes approach to inferring large-scale gene association networks." Bioinformatics **21**(6): 754-764.

See, H. T., J. J. Kavanagh, et al. (2003). "Targeted therapy for epithelial ovarian cancer: current status and future prospects." Int J Gynecol Cancer **13**(6): 701-734.

Shahab, S. W., L. V. Matyunina, et al. (2012). "The effects of MicroRNA transfections on global patterns of gene expression in ovarian cancer cells are functionally coordinated." Bmc Medical Genomics **5**: 33.

Shahab, S. W., L. V. Matyunina, et al. (2011). "Evidence for the Complexity of MicroRNA-Mediated Regulation in Ovarian Cancer: A Systems Approach." PLoS ONE **6**(7).

Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-2504.

Siciliano, V., I. Garzilli, et al. (2013). "MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise." Nat Commun **4**: 2364.

Singer, G., R. Oldt, 3rd, et al. (2003). "Mutations in BRAF and KRAS characterize the development of low-grade ovarian serous carcinoma." Journal of the National Cancer Institute **95**(6): 484-486.

Soneson, C., H. Lilljebjorn, et al. (2010). "Integrative analysis of gene expression and copy number alterations using canonical correlation analysis." BMC Bioinformatics **11**: 191.

Stephens, P. J., P. S. Tarpey, et al. (2012). "The landscape of cancer genes and mutational processes in breast cancer." Nature **486**(7403): 400-404.

Su, S., Q. Liu, et al. (2014). "A positive feedback loop between mesenchymal-like cancer cells and macrophages is essential to breast cancer metastasis." Cancer Cell **25**(5): 605-620.

Tan, P.-N., M. Steinbach, et al. (2005). Introduction to Data Mining, Addison-Wesley.

Thompson, M. R., D. Xu, et al. (2009). "ATF3 transcription factor and its emerging roles in immunity and cancer." J Mol Med (Berl) **87**(11): 1053-1060.

Tijsen, A. J., Y. M. Pinto, et al. (2012). "Circulating microRNAs as diagnostic biomarkers for cardiovascular diseases." Am J Physiol Heart Circ Physiol **303**(9): H1085-1095.

Tokunaga, H., Y. Takebayashi, et al. (2008). "Clinicopathological significance of circadian rhythm-related gene expression levels in patients with epithelial ovarian cancer." Acta Obstet Gynecol Scand **87**(10): 1060-1070.

Tsang, J., J. Zhu, et al. (2007). "MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals." Molecular Cell **26**(5): 753-767.

Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-5121.

Valbuena, A., S. Castro-Obregon, et al. (2011). "Downregulation of VRK1 by p53 in response to DNA damage is mediated by the autophagic pathway." PLoS One **6**(2): e17320.

Vaughan, A. E., R. Mendoza, et al. (2012). "Xpr1 is an atypical G-protein-coupled receptor that mediates xenotropic and polytropic murine retrovirus neurotoxicity." Journal of Virology **86**(3): 1661-1669.

Vera, J., X. Lai, et al. (2013). "MicroRNA-regulated networks: the perfect storm for classical molecular biology, the ideal scenario for systems biology." Adv Exp Med Biol **774**: 55-76.

Vidal, M., M. E. Cusick, et al. (2011). "Interactome networks and human disease." Cell **144**(6): 986-998.

Wang, X. (2014). "Composition of seed sequence is a major determinant of microRNA targeting patterns." Bioinformatics **30**(10): 1377-1383.

Welsh, J. B., L. M. Sapinoso, et al. (2001). "Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer." Cancer Research **61**(16): 5974-5978.

Witkos, T. M., E. Koscianska, et al. (2011). "Practical Aspects of microRNA Target Prediction." Curr Mol Med **11**(2): 93-109.

Wolfram Research, I. (2010). Mathematica. Champaign, Illinois, Wolfram Research, Inc. . **Version 8.0**.

Wrzeszczynski, K. O., V. Varadan, et al. (2011). "Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer." PLoS ONE **6**(12): e28503.

Wu, C., J. Zhu, et al. (2012). "Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes." BMC Bioinformatics **13**: 182.

Yano, M., M. Ouchida, et al. (2004). "Tumor-specific exon creation of the HELLS/SMARCA6 gene in non-small cell lung cancer." International Journal of Cancer **112**(1): 8-13.

Yates, L. R. and P. J. Campbell (2012). "Evolution of the cancer genome." Nat Rev Genet **13**(11): 795-806.

Ying, H., J. Lv, et al. (2013). "Gene-gene interaction network analysis of ovarian cancer using TCGA data." J Ovarian Res **6**(1): 88.

Ying, L., D. Su, et al. (2013). "Genotyping of stathmin and its association with clinical factors and survival in patients with ovarian cancer." Oncol Lett **5**(4): 1315-1320.

Zhao, M., J. Sun, et al. (2012). "Distinct and competitive regulatory patterns of tumor suppressor genes and oncogenes in ovarian cancer." PLoS One **7**(8): e44175.